

November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases; older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New).

Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

this Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2006, 20:42:45 ; Search time 625 Seconds
(without alignments) 5731.666 Million cell updates/sec

Title: US-10-017-471B-19
Perfect score: 4346
Sequence: 1 qtcgacgcagcgcgtcggtt.....gcctagagccgcgcgggtacc 4346

Scoring table: IDENTITY NUC

scoring cable: IDENTIFIER=NOO
Gapop 10.0 , Gapext 1.0

Searched: 6049916 seqs, 412136615 residues

Total number of hits satisfying chosen parameters: 12099832

Minimum DB seq length: 0

Minimum	DB	seq	length:	0
Maximum	DB	seq	length:	20

Post-processing: Minimum Match 0%

POST-PROCESSING: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : Published Applications_NA_New.*
1: /cgm2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
2: /cgm2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
3: /cgm2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
4: /cgm2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
5: /cgm2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
6: /cgm2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
7: /cgm2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
8: /cgm2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
9: /cgm2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
10: /cgm2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
11: /cgm2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
12: /cgm2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
C 1	82.6	1.9	10968	8	US-11-075-185-35	Sequence 35, Appl
C 2	82.6	1.9	78869	8	US-11-075-185-1	Sequence 1, Appl
C 3	78	1.8	846	7	US-10-467-657-1341	Sequence 1341, Ap
C 4	78	1.8	5706	8	US-11-052-554A-519	Sequence 519, App
C 5	74.6	1.7	88421	8	US-11-025-109-1	Sequence 1, Appl
C 6	72.5	1.7	2406	8	US-11-052-554A-550	Sequence 550, App
C 7	72	1.7	3990	8	US-11-052-554A-520	Sequence 520, App
C 8	71.8	1.7	5706	8	US-11-052-554A-519	Sequence 519, App
C 9	71.4	1.6	88421	8	US-11-025-109-1	Sequence 1, Appl
C 10	70.8	1.6	2196	8	US-11-052-554A-539	Sequence 539, App
C 11	69.8	1.6	5679	8	US-11-075-185-36	Sequence 36, Appl
C 12	69.8	1.6	78869	8	US-11-075-185-1	Sequence 1, Appl
C 13	66.6	1.5	5679	8	US-11-075-185-36	Sequence 36, Appl
C 14	66.6	1.5	11070	8	US-11-075-185-34	Sequence 34, Appl
C 15	66	1.5	2250	8	US-11-052-554A-532	Sequence 532, App
C 16	64.6	1.5	11070	8	US-11-075-185-34	Sequence 34, Appl
C 17	64.4	1.5	989	8	US-11-124-368A-159	Sequence 159, App
C 18	64.4	1.5	1980	8	US-11-129-143-177	Sequence 177, App
C 19	64.2	1.5	4617	8	US-11-052-554A-530	Sequence 530, App
C 20	63.6	1.5	3240	8	US-11-052-554A-529	Sequence 529, App
C 21	63.4	1.5	882	8	US-11-128-061-619	Sequence 619, App
C 22	63.4	1.5	882	8	US-11-128-061-4261	Sequence 4261, Ap

C	23	63.4	1.5	882	8	US-11-128-049-619	Sequence 619, App
C	24	63.4	1.5	882	8	US-11-128-049-6261	Sequence 4261, App
C	25	62.8	1.5	958	7	US-10-775-169-118	Sequence 118, App
C	26	62.8	1.4	8651	7	US-10-432-483-48	Sequence 48, Appl
C	27	62.6	1.4	1082	8	US-11-128-061-685	Sequence 685, App
C	28	62.6	1.4	1082	8	US-11-128-049-685	Sequence 685, App
C	29	62.4	1.4	2515	8	US-11-136-537-2896	Sequence 2896, App
C	30	62	1.4	2562	8	US-11-052-554A-533	Sequence 533, App
C	31	61.6	1.4	10968	8	US-11-075-185-35	Sequence 35, Appl
C	32	61.2	1.4	2302	8	US-11-110-082-30	Sequence 20, Appl
C	33	61.2	1.4	10524	8	US-11-075-185-37	Sequence 37, Appl
C	34	61	1.4	1635	7	US-10-432-483-16	Sequence 16, Appl
C	35	61	1.4	8651	7	US-10-432-483-48	Sequence 48, Appl
C	36	60.8	1.4	1263	7	US-10-858-730-149	Sequence 149, App
C	37	60.2	1.4	2304	8	US-11-052-554A-538	Sequence 538, App
C	38	60	1.4	3116	8	US-11-088-785-3	Sequence 3, Appl
C	39	59.8	1.4	2514	8	US-11-052-554A-543	Sequence 543, App
C	40	59.2	1.4	3408	7	US-10-858-730-40	Sequence 40, Appl
C	41	57.8	1.3	2562	8	US-11-052-554A-533	Sequence 533, App
C	42	57.8	1.3	3240	8	US-11-052-554A-529	Sequence 529, App
C	43	57.8	1.3	14172	8	US-11-075-185-2	Sequence 2, Appl
C	44	57.6	1.3	1086	8	US-11-052-554A-553	Sequence 553, App
C	45	57.4	1.3	1731	8	US-11-052-554A-548	Sequence 548, App

ALIGNMENTS

```

RESULT 1
US-11-075-185-35/c
; Sequence 35, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075.185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35
; LENGTH: 10968
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-11-075-185-35

```

	Query Match	1.9%	Score 82.6;	DB 8;	Length 10968;
	Best Local Similarity	43.1%;	Pred. No. 0.00012;		
	Matches 624;	Conservative 0;	Mismatches 809;	Indels 16;	Gaps 4;
Qy	3	CGACGACGGCGTTCGGGTTTCGACGCCGACGGGTACTCGTTCGGGGCACCGGGCACCGGG	62		
Db	4138	CGTCCAGCGCTCGGGCCCGAGGTCGACGAGGCGCACCGCCGGCGGGGTTTCTCTGGC	4079		
Qy	63	TCCTGGCTCGATGACCGACCGCATCGAGGACGTCGGCTGGCGGCTCCTGATGTAGACGG	122		
Db	4078	GCGCCGTGCGCACAAGCCCCACAGCGCGCGGTGGGTCAACGCCGCGACGCCGTCGTCTGG	4019		
Qy	123	CCCGCGCGGGGACGCGACATCGACGTCCATCTCCCACTCGCGCCCCGGGAAAGTAGACAC	182		
Db	4018	GCCCCGGGGCCACAGCGCGCGGCTCACCACACACAGCTCGCTGTCCGCGAGCGCGCCT	3959		
Qy	183	CGCACCCGCGACGTGACGCCATGGGAGGGCCAGCTCGCGCGACCGATCAACCCCTGGGCTTC	242		
Db	3958	CCGAGACCACTCTGCACACGACGACGAGGACCCGCTCGCTCGCGCTGCGCGCCGCCGCA	3899		
Qy	243	GGCCGAAGGCTTCGGGTGTTCGGCCGCCACGATGCGGAACGGCTCGCGGGCCGGCCCG	302		

Db 3898 CAGCGCGTCCCGGGTTCGCGCGCGCGGTGATCAGACAGCGCGCGCGCGCG 3839
Qy 303 TCACCTCCCGCAGCGTCCGCGACGACTGCGCTTTGGCCCCCGCCCGCTGCGCGC 362
Db 3838 CGCCTCCGCGAGCGCGCGGATGAGTCCGCGAGCGCGCGCTGCGCGCGCGCGCG 3779
Qy 363 TCTCAGGAACGCGTCCGCGAGCGGCGCTCGGATATCTTCGGAGGATCTGCTCGCGCGCG 422
Db 3778 CCGGAGCGCCCGTCCGCGCGGACGATGTGCTCTCCGCGCGCCACGC---CGCTT 3722
Qy 423 CTTCCAGCGCCAGCTCTCTGAAACGCGAGTACCGCCCGACACCTCGAGAGCGCGCG 482
Db 3721 CTTCCAGCGCCAGCTCTCTGACAGTACCGGTCAGGTACGTCTGCTGCGACCGCGCG 3662
Qy 483 TCTGCTGCTGCGCGGTGATCGCTCGGCGCCGCGACAGCGGTACGCGCTCCCTGTTGGC 542
Db 3661 CAGCTTTGCGCAGCTGCTCGCGCTGCGCCACCGCGCGCGGAAAGTCTCTCGCGCGCGA 3602
Qy 543 CGGCGCGGTGAGTCCCTGACCGCACTTCGCGATGTGCGCGGATCGAGCGAGGCA 602
Db 3601 CCGGCTGCTGCGCGTCCGCGCGAGTCTGACGCGCGCGAGTGTCTCCATCGTGTG 3542
Qy 603 CCGGCGAGTGTGCGGTACAGCGCGGACACGCGTCCGACCGGATGCGCGCGCGCCAG- 661
Db 3541 CGTCTCGACCTCCAGCGCGCGCGCGCGCGTGTGCCCCCGCGCTCGAGCGCGCAGT 3482
Qy 662 --GACAGCGTGTGACATGAAGTCTTGGCCCGCGAGGAAGGTCCAGTCTAGCCCGGACT 719
Db 3481 CCGACCAACGGAAGGCGATCGGATGCTCCCGCGCGAGCGCTGCGCGGAGTCCG 3422
Qy 720 CGGTAAGCGCGGTCTGCTGCGCTGCGCGCGGTGATGAATGCTGCGCGCGCGGT 779
Db 3421 CCGCAGGAGGCGCTGAGCGCGCGCATCGAGCAAGCGCGGATCGAAGCGGAAGTCTGTG 3362
Qy 780 CCCCACCGCGAGCATGACAGCTTACAGGTGCGGACCGCGCTCGCGCGCGCGCG 839
Db 3361 CCGGCTCCGAGACCCCTCGGGCAACGCGAGTCCGCGTGAATGACCGCGCCCGCGCG 3302
Qy 840 CCGGAAACGCTGCTGCTCGGCTCGGTGCGACTGTTCTGTAAGAGGAACCGCGCGCA 899
Db 3301 AGCGCGCGTACGCGCGCGGAAAGCGCGCGCGACCAAGCGTCCGAGCGCTCGACCCACC 3242
Qy 900 CCGGTTGAGGCGCGGTTCAGGCGCGCGCGGTGCGGCTACTCGCGCGCGCGAGCTCGA 959
Db 3241 GGTGACGTGACACCGGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3187
Qy 960 GGTTCGGGCGGTGACGCTCACCGCTTCGCGCGCGCGCGCGGAGGACTCTGACGGGACCGG 1019
Db 3186 GGTGTCGCGCGCGCGCGGTGCGCCAGACGCGCGCGCGGTGCGCGTCCAGCGCGC 3127
Qy 1020 TCCGGGCGAGAGTGGGCGACCTGACCGCGCGACACACCGGTTCAGCGCGGTTCACAGAA 1079
Db 3126 GTGTCGCGCGGTGCGGCGCGCTGTGACCGGAAAGCGCGCGCGCGCGCGCGCGCG 3067
Qy 1080 TCACCTCGGGGCTCTCTCGGCGAGCGAGCGAGGCGCGCTCGAATACATATGAGGGG 1139
Db 3066 GCGCTTCAACCGTACCTGACGCGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 3007
Qy 1140 AAGGCGAGGATCTGCGCGCGCGCGGAAACCGCGGATGTTGCGCGCGCGCGCGCGGT 1199
Db 3006 GATGTCGAGTCTGACACGAGGCGAGTCCGAGCGCATCGCGCGCGCGAGCGCGCGT- 2948
Qy 1200 CAGCGGAGAACGCGGCGCGGACAGCGTGTGTGAGGAAGACGAGGTCGCGTCTGTATGC 1259
Db 2947 ---CGAGAAACGCGTGTGCGGAGGATCACCGTGTGCGGACCAACGTCGTGTCAGGACC 2892
Qy 1260 CCGGTGACCGGCGAGTGTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1319
Db 2891 CACGCGTGTCTACACCGTTCAGACGCGTTCGCTCAGACGTAAGTGTGCGGTCGCGCAGC 2832
Qy 1320 CAGCACGCGGCTGTGAACTCTCGCGTACCGGAGGAAACCGGTCGCGCGGATCGACGGCAGC 1379

Db 2831 CGACCGCGCGGTGAGCCACGCGGTCTTCAACGCGAGGAGCCCTCCGCTCTCCAGCCCT 2772
Qy 1380 AAGGCGCGCGAGCGGTCAGAGGACCGCGCTGCGCGCTCGAGCAGCAGCATG 1439
Db 2771 GCGCGCGCGCGCGCGCTCGAGCCAGTACGCTGCGCGTGGAAACGCGTAGGTGCGGCAAG 2712
Qy 1440 CCGGTACG 1448
Db 2711 TCCGCAAG 2703

RESULT 2
US-11-075-185-1/c
; Sequence 1, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; PRIOR FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08
; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 78869
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-11-075-185-1

Query Match 1.9%; Score 82.6; DB 8; Length 78869;
Best Local Similarity 43.1%; Pred. No. 0.00014;
Matches 624; Conservative 0; Mismatches 809; Indels 16; Gaps 4;

Qy 3 CGACGAGCGGTTCGGTTCGAGCGCGGCGGACTGTTCCCGGCGACCGGCGACCGGG 62
Db 19405 GGTCCAGCGGTTCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19346
Qy 63 TCTCGGCTCGATGACCGGACCGCATCGAGGAGCTCGGCTGGCGGCTCTGATAGTACG 122
Db 19345 GCGCGGTGCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19286
Qy 123 CCG 182
Db 19285 GCG 19226
Qy 183 CGACCGCGGAGTGAAGCGGATGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 242
Db 19225 CCGAGAGCG 19166
Qy 243 GCGCGAGGCTTCGCGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 302
Db 19165 CGAGCGGTTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19106
Qy 303 TCACCTCCCGCAGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 362
Db 19105 CGCGCTCCG 19046
Qy 363 TCTCAGGAACGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 422
Db 19045 CCGGAGCG 18989
Qy 423 CTTCCAGCG 482
Db 18988 CTTCCAGCG 18929
Qy 483 TCTGCTGCTGCGCGGTGATCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 542

Db	1928	CAGCTTGGCGCAGCTGCTCCGCGTCCGCCACCGCGCCCGGAACTCCTCCGTGGCGCGGCA	18869
Qy	543	CGGCGCGGGTCAGTGCCTGACCGCCACTTCGCGCATGTGCGCGCGGATCGACGAGGCAA	602
Db	19868	CCGGCTGCCCTCGCGTCCGCGAGCTCGATCGACGCGCGAGCTGGTCCCACATCGCTGG	18809
Qy	603	CCGGGACGTCGCGTACAGGCGCGCGAACACGCGCTCGGACCGGATGGCGGGCGCCAG-	661
Db	18808	CGTCTCTGCACTCCAGCGCGCGCGGAGCGCGCTGCCCGCGCCCTCGAGCGCCACGT	18749
Qy	662	--GACACGCTGTTGGACATGAAGTCTTGCGCCGCGAGGAAGTCTCAGTCTTAGCCCGGACT	719
Db	18748	CCGACACGCGAAGGCGATCGGATGCCCTCCGCGCGACAGAGCTTGGCGCGAGCTCCG	18689
Qy	720	CGGTAACGCCCCCTCGTCTTCGCGCTGCGCGCGGTGATGAAGTCTGTCGCGGCCCGTT	779
Db	18688	CCCGAGGAGGCGCTTGAGCGCGCATCGAGCAAGCGCGGATGCAAGCCGAAGTCTTCG	18629
Qy	780	CCCCACCGCGAGCATGGACAGCTTACCAGGTGCGGACGCGCGCTCGCGCGCGCGCG	839
Db	18628	CCCGCTCCGAGACCCCTCGGGCAAGCCAGGTTCGGCTAGACATCCGCCCCACGCGCC	18569
Qy	840	CCGCGAAACCTCGTTCGTCGCGTCCGGTGGCATGTTTCGTGACGAGGAAACGCGCCCGCA	899
Db	18568	ACGCGCGCTGACGCCCCCGGAACCGCGGCGGACCCACGCTCGAGCGCTCGACCCACC	18509
Qy	900	CCCGTTGAGGGCCCGGTCCAGGCCCGGGCGGTGCGCTGTCGCGCGCGACAGCTCGA	959
Db	18508	GCGTGACGTGACCGGCTCGGCCCCC-----GCCGCGGCCACAGCGCGCGCCGCGCCACGC	18454
Qy	960	CGTTCGGGCGGTGACGCTCACCCGTTCCGCGCGCGCGGCGAGGACTCTGACGGGACCG	1019
Db	18453	CGTCTGTCGCGCGCGCGCGCGCTCGGCCAGCACGCGCGCGCGTGGCGCGTCCACGGCGC	18394
Qy	1020	TCCGGGCCACAGCTGGGCGACCTGACGCGCCGACACACCGGTCACGCGGTCAAGAA	1079
Db	18393	GTGTCGCGCGCTGGGCGCGGTGTGGACCGCGAGCCCGCGCGCGTCCGCTCCAA	18334
Qy	1080	TCACTCGGGCTCTCTCGGGCAGCGAGGAGGGGCGCTCCGAAACATATAGAGGG	1139
Db	18333	CGCTCCACCGTCACTTCGACGCTCACCGCCCGCGCGCGGCGAGCGCGCGCTC	18274
Qy	1140	AAGGCAAGATCTGCCCCGGGGCGGAAACGGGCGATGTTTCGCGCGCCCGGGCGGTGTT	1199
Db	18273	GATCGTAGCTCTGACACCGAGGGCAGTCCGACGGCATCGGCGCGCGGAGCGCCAGCT-	18215
Qy	1200	CAGCCGAGAAACCGGGCGCGACAGCGTGGTGAGGAACAGAGGTGCGCTCTGATGC	1259
Db	18214	----CGACGAAACGCGTGCCTCGGGAGGATCAACGCTGCCACGACCACTGGTCCAGGACC	18159
Qy	1260	CCGCTGACCCGACAGTGTGTCAGCCCGCGCGCGCCCGGGCGGACCGTCCGCTGGATC	1319
Db	18158	CACGCTGCTCGAACCGTTCGACAGCCGCTCCGCTCAGCACGTAGCCGCTCGCGTTCGGCAGC	18099
Qy	1320	CAGCAGGGCTGTGCAACTCCGGGTACCGGACGAAACCGGGTGC CGCCGATCGACGGCAGC	1379
Db	18098	CGCACCCCGCGCTGAGCCACAGGGTGTTCACCGCCAGAGAGCCCTCCGCTCTCCAGCCCT	18039
Qy	1380	AAGGGCCCGCGAACCGGTTCAGAGGCGACGCGCTGCGCTGCGCTTCGAGACGACGATG	1439
Db	18038	GCCCGCGCGCGCGCTTCGAGCCAGTAGCGCTGCGCTGGAAACGCGTAGTTCGGCAG	17979
Qy	1440	CCCGGTACG	1448
Db	17978	TCGCGCACG	17970

RESULT 4

```

RESULTS - 4
US-11-052-554A-519/c
; Sequence 519, Application US/11052554A
; Publication No. US20050289866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD
; TITLE OF INVENTION: PROTEINS OF THERM
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589, 2
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/

```

RESIT.T 3

US-10-467-657-1341/c
; Sequence 1341, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA


```

; LOCATION: (9464)..(8130)
; OTHER INFORMATION: ORF 6; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9691)..(10761)
; OTHER INFORMATION: ORF 7; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (12751)..(10829)
; OTHER INFORMATION: ORF 8; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (13617)..(12802)
; OTHER INFORMATION: ORF 9; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15203)..(13614)
; OTHER INFORMATION: ORF 10; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15591)..(15863)
; OTHER INFORMATION: ORF 11; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (15880)..(19035)
; OTHER INFORMATION: ORF 12; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (19032)..(39713)
; OTHER INFORMATION: ORF 13; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (39713)..(65800)
; OTHER INFORMATION: ORF 14; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (65826)..(66530)
; OTHER INFORMATION: ORF 15; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (66546)..(67370)
; OTHER INFORMATION: ORF 16; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (67384)..(70059)
; OTHER INFORMATION: ORF 17; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (70099)..(70662)
; OTHER INFORMATION: ORF 18; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (70659)..(71906)
; OTHER INFORMATION: ORF 19; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (73439)..(71964)
; OTHER INFORMATION: ORF 20; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (74216)..(73563)
; OTHER INFORMATION: ORF 21; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (75424)..(74213)
; OTHER INFORMATION: ORF 22; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (75535)..(76464)
; OTHER INFORMATION: ORF 23; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (78110)..(76449)

```

```

; OTHER INFORMATION: ORF 24; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (79864)..(78107)
; OTHER INFORMATION: ORF 25; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (81624)..(79861)
; OTHER INFORMATION: ORF 26; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (81909)..(81682)
; OTHER INFORMATION: ORF 27; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (82346)..(82062)
; OTHER INFORMATION: ORF 28; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (82587)..(84446)
; OTHER INFORMATION: ORF 29; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84481)..(85548)
; OTHER INFORMATION: ORF 30; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (85556)..(86845)
; OTHER INFORMATION: ORF 31; positive strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (87372)..(86803)
; OTHER INFORMATION: ORF 32; negative strandedness
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (87494)..(88420)
; OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
; US-11-205-109-1

```

```

Query Match      1.7%; Score 74.6; DB 8; Length 88421;
Best Local Similarity 48.7%; Pred. No. 0.0025;
Matches 320; Conservative 0; Mismatches 304; Indels 33; Gaps 3;

QY 3152 CGGCCAGTCCGCGCCCTCGGGGGTCTGCGCGCGCGTTCATGTCGCTCTCCA 3211
DB 67206 CGTCGAGCGACCGGCGGATCTCCGACAGCATCCGCGCCCTCGAACTGCGGATCTCGCGCA 67147
QY 3212 CGAAACCGCGCGCACCGCGTTGACCGGTATCCCGCTTCCCGCAGTTGCCCTGGCCAGG 3271
DB 67146 CGAAACCGCGCGACGCGGTTGACCCCGACCTTCGCTGGCGCGTAGTCGAGCGGCGG 67087
QY 3272 CGAGCGTGAGCGTGTCCACCGCACCCCTTGGTTCATCGCGATATCCGATGGAATCGGGAAACG 3331
DB 67086 CCCGGGTGAGCCCGACGATGCGTCTTGGCGGACGATAGCCGCGAAGTCCCGTAGC 67027
QY 3332 CGCGCCGGTCCGCGCACGAGATGTTGATGATCCGCCCGCGCGTCCGCGCATGCTTTCA 3391
DB 67026 CGACCTGACCGCGCGTCCGAGCGATGTTGACGATCTGCGGAGCGGCGCTCGGTCACTG 66967
QY 3392 GTCCGTGCTGGAC-----CAGGAACAGCGGTGCGCGGACGTTGACGGCGACGATCGGT 3445
DB 66966 GCGGCGCAACCGCGCCCGTTCATGCGCCACGCGCGGAGAGGTCGACGTCGATCATCAGT 66907
QY 3446 CGAAGACCTCTCGGTGACTTCCGTGATCCGTCGCGAGCGGTGACGCCCGCGGTGTTTCA 3505
DB 66906 GCCACTCGTCTCGGTGATCTGTTGGTGA---CTTTCCGAGCGCGCGCGATGCCCG 66850
QY 3506 CCAGGATGTCGAATCCGGCGGCACTCCGAACTCCGCATCCCGGCGTCCGAACGCGCGT 3565
DB 66849 CGTTGTTGACGAGCACGTCGATCGCGCGCGAAGCGGTGATCGCCCGATCGCGCGCGG 66790
QY 3566 AGAGCGCGCGCGTCACTCCAGCAGCGCGAGTTCGCGCCCGGATGCCCAACGCTGTCGCG 3625

```


; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 520
; LENGTH: 3990
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-520

Query Match 1.7%; Score 72; DB 8; Length 3990;
Best Local Similarity 42.7%; Pred. No. 0.0052;
Matches 369; Conservative 0; Mismatches 495; Indels 0; Gaps 0;
QY 3483 GCCGCTGACGCCCGCGTGTGTTCCACGAGATGTGCAACTCGGCGCGCACTCCGAACTCGCC 3542
DB 1893 GCCGCGCCCGCGCGCCCGCGGAGGTGCGTTCGCGCGGTTCGCGCGGTGCGCGC 1834
QY 3543 CATCCCGGCGTGAACCGCGGTAGAGCGCGCGGTCAACCAAGCGCGAGTTCGCGC 3602
DB 1833 GTTGCCTGCGTGTGCGGCGCGCGCGCGCGTCTGACCGCGGTGATCCCGC 1774
QY 3603 CGGATGCGCAACGCTGTCCGCGCTGCTCGGATGCTCTGACGCTCTCTCGCGCGC 3662
DB 1773 GCCGCGCATCCACAGACACCGCATGCGCGGTACCGCGGTTCGCGCGCTTCCACC 1714
QY 3663 CGCCTCGCTGCTGCGTGTGCACTGCAACGAGCGCGCGTTCGCGCGCGCGCGCGC 3722
DB 1713 AACTCATCGCGCGGTATGCAAGTGTGCGTTCGCGCGGTTCGCGCGGTTCGCGC 1654
QY 3723 GATACCGGTTCGATGCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGG 3782
DB 1653 GCCGCGCGCGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTCG 1594
QY 3783 TCTTCATACCTGCTCCATGTCAGCATATACGCGCGCGCGCGCGCGCGCGCGCGCG 3842
DB 1593 GGTCTGAAGCGCTTACCGCGCTTGCACCGCGCTCCACCGCGCGCGCGCGCGCGT 1534
QY 3843 GCGCGCGCTGCGCGTTCGAACTGCAAGCTCACAGCTTACCTGACCGCGTTCAGCGG 3902
DB 1533 GCGCGCTTGCCTGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCG 1474
QY 3903 GCGGAGTGGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTC 3962
DB 1473 GCGGTGGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTC 1414
QY 3963 GCGGTGAGGCTCAGGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCG 4022
DB 1413 GCTGCTTGGGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTC 1354
QY 4023 GAGGTGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTC 4082
DB 1353 GCGGTGAGGCTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTC 1294
QY 4083 CCAGAGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTC 4142
DB 1293 GCGCGCGCTTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTC 1234
QY 4143 GCGCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTC 4202
DB 1233 GCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTC 1174
QY 4203 AGCGCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTC 4262
DB 1173 GCTGGCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTC 1114
QY 4263 CCGCGAGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTCGCGCGTTC 4322

DB 1113 CCCGCGCGCGCGCGGTTCACCGCGTTCACCGCGCAATCAAGGCGCGGAAATCC 1054
QY 4323 CGACGCGTAGAGCGCGCGGTAC 4346
DB 1053 GCGCGCGCGCGCGCGGTCCC 1030
RESULT 8
US-11-052-554A-519
; Sequence 519, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 519
; LENGTH: 5706
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-519

Query Match 1.7%; Score 71.8; DB 8; Length 5706;
Best Local Similarity 42.2%; Pred. No. 0.0057;
Matches 606; Conservative 0; Mismatches 817; Indels 14; Gaps 3;
QY 2909 CCGCGAAGCGCGCGCGCGGTACCGACCGCGCGTTCGCGCGCGCGCGCGCGCGCG 2968
DB 3453 CCGCGCGAAGCGCGCGCGGTGCGGGAACCTTCGGCGGTTCGCGCGCGCGCGCG 3512
QY 2969 CTACGCGCGCGCGCGCGCGGTAGTCTGCGTTCGCGTTCGCGTTCGCGTTCGCG 3028
DB 3513 CCGCGCGAAGCGCGCGCGGTGCGGCGGTTCGCGCGCGCGCGCGCGCGCG 3572
QY 3029 GAATCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCG 3088
DB 3573 TAAGCGCGCGCGCGCGCGGTGCGGGAACCTTCGGCGGTTCGCGCGCGCGCGCG 3632
QY 3089 CCAGAAAGCGCGCGCGGTGCGGCGGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCG 3148
DB 3633 TCTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCG 3692
QY 3149 TGGCGCGCGTTCGCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGTTCGCGTTCGCG 3208
DB 3693 CAGCGCGCGCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTC 3745
QY 3209 CCAGAAAGCGCGCGCGGTGCGGCGGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCG 3268
DB 3746 GGGCGCGGTTCGCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 3805
QY 3269 GGGCGCGGTTCGCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 3327
DB 3806 GCGAAGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 3865
QY 3328 AACGCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 3387
DB 3866 ACGCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 3925
QY 3388 TTCACTTCGCGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 3447
DB 3926 GCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGTTCGCGCGGT 3985
QY 3448 AAGACCTTCGCGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGTTCGCGGT 3507

Db 3986 ACGGCGGCAAGTGGTGTCCGGCGGCGCACCGGGCCCAAGGCGGCAAGCGCCCTGGAACA 4045
 Qy 3508 AGGATGTCGAATCTGGGCGGCACTCCGAACTTCGCCATTCCTGGCGGTTCGAACCGCCGGTAG 3567
 Db 4046 GCACCGGCTAGCCACGAGCCCGCAGCGGTGACCGCGGCAACCGCGGGCGCGCGGGCGG 4105
 Qy 3568 AGCGGGCGGCGTACACGACGCGAGTTTCGGCGCGGATGGCCAAACCGCTGTCCGCGG 3627
 Db 4106 GCGGCAACCGGCGGACCGCGAGCGCTTTATTCGGGGTTCGGCGGACCGCGGGGAGACG 4165
 Qy 3628 CTGCTCCGGATGCTCTGACAGCTCTCTCGCGCGCGCTCTGCTGCTCTCCCTAGTGGACT 3687
 Db 4166 GCGGCAACCGGCGGCTTCGGCGGCTTGGCAACACCGGCGGAAACCGGGCAACCGCGTA 4225
 Qy 3688 GCCACGAGCGCCCTTCGCGGCGCACCGCGAGGCGATACCGCGTTCGATCCCGCGGCTT 3747
 Db 4226 TCGGCGGCTCTCC 4285
 Qy 3748 CCCCGGTACAGGGCGGTCTTGCCTTCAGCGGTCTTCATACCTGCTCCCATGTGCA 3807
 Db 4286 AAGACGGCAACGGCTTCG 4345
 Qy 3808 CGCATATCAGCCCGCGCTGCTGAGCGACCCATGCGCGCGCTTCGCGCGCTTCGAATCG 3867
 Db 4346 GCGCGCGGCGATCAACGTGTGAG 4405
 Qy 3868 ACGGTACAGCGCTTACTGTGACCGCGTTCAGACGCGCGCGCGCGCGCGCGCGCGCGT 3927
 Db 4406 GCCAGAACGGAACGACGCGGTGCGGAGCGGGCGCGCGCGCGCGCGCGCGCGCGCGTACG 4465
 Qy 3928 GGGGCGATCGGCG 3987
 Db 4466 GCGGCAAGCGGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTTC 4525
 Qy 3988 GGACCGCGCGCGCGTTCAG-----GGCACCGACCGGATCGAGTTCGGCGGTGCGCACGC 4041
 Db 4526 GCGACGGCGGATCGCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4585
 Qy 4042 GGCCACGATCTCTCG 4101
 Db 4586 GGGACCGCA 4645
 Qy 4102 AGTCTGAGTGCAGTGGCG 4161
 Db 4646 AAGCGCGCTTGAACAGCACCGCGCGTACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 4705
 Qy 4162 GGGAGCTCGTCC 4221
 Db 4706 GGGCG 4765
 Qy 4222 CAGGCGTCTCTCAGGACGCTTAGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4281
 Db 4766 GCACCG 4825
 Qy 4282 CCGCGCGTCTCGCGTTCG 4338
 Db 4826 CCGGCAACCGCGGTATCG 4882

RESULT 9

US-11-205-109-1
 ; Sequence 1, Application US/11205109
 ; Publication No. US20050287641A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Parnet, Chris
 ; APPLICANT: Zazopoulos, Emmanuel
 ; APPLICANT: Staffa, Alfredo
 ; TITLE OF INVENTION: GENE CLUSTER FOR RAMOPLANIN BIOSYNTHESIS
 ; FILE REFERENCE: 3002-2US
 ; CURRENT APPLICATION NUMBER: US/11/205,109
 ; CURRENT FILING DATE: 2005-08-17
 ; PRIOR APPLICATION NUMBER: US/09/976,059
 ; PRIOR FILING DATE: 2001-10-15

; PRIOR APPLICATION NUMBER: US 60/239,924
 ; PRIOR FILING DATE: 2000-10-13
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 88421
 ; TYPE: DNA
 ; ORGANISM: Actinoplanes sp.
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (2077)..(3078)
 ; OTHER INFORMATION: ORF 1; positive strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (3118)..(4032)
 ; OTHER INFORMATION: ORF 2; positive strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (4038)..(5048)
 ; OTHER INFORMATION: ORF 3; positive strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (6665)..(5814)
 ; OTHER INFORMATION: ORF 4; negative strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (7703)..(6693)
 ; OTHER INFORMATION: ORF 5; negative strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (9464)..(8130)
 ; OTHER INFORMATION: ORF 6; negative strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (9691)..(10761)
 ; OTHER INFORMATION: ORF 7; positive strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (12751)..(10829)
 ; OTHER INFORMATION: ORF 8; negative strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (13617)..(12802)
 ; OTHER INFORMATION: ORF 9; negative strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (15203)..(13614)
 ; OTHER INFORMATION: ORF 10; negative strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (15591)..(15863)
 ; OTHER INFORMATION: ORF 11; positive strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (15880)..(19035)
 ; OTHER INFORMATION: ORF 12; positive strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (19032)..(39713)
 ; OTHER INFORMATION: ORF 13; positive strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (39713)..(65800)
 ; OTHER INFORMATION: ORF 14; positive strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (65826)..(66530)
 ; OTHER INFORMATION: ORF 15; positive strandedness
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (66546)..(67370)
 ; OTHER INFORMATION: ORF 16; positive strandedness
 ; FEATURE:

Db	40915	ACGCCCGCGCGTCTCACCGGGGCACTGATCGCGCGCGCGCGCTGTTCGACGCGCGT	40974
Qy	3784	CTTCCATACCTCGTCCCATGTGCAGCATATCAGCCCGCGCGTGCGTGAGGACCCCATG	3843
Db	40975	GGTGGAGCGGCTGGCCGCGCTGCTGGCCCGCGCGCTCGAGGGCGTGGCGCGGACCCGCG	41034
Qy	3844	GCGGCGCTCGGCGGTTGGAATCGACGGTTCACAGCTACTCTGTGACCGCGTCAGACGGG	3903
Db	41035	CACGCGGCGCGGCACTCGACTGTCTCTCCCGGGGACCGCCCGGTGATCTTGGCGG	41094
Qy	3904	CCGAGTGGGCCCGGTTGGACGGTGGGGCCAGATCGGCGCGCGCACCGGGAAACCGCGC	3963
Db	41095	CTGGAACGACACCGCGCTCCGCGCGCGCGGACTGGTGGCGGACCTG---TTGCGCG	41150
Qy	3964	CCGCTCAGGGGTCAAGGGGTCCCGGGACCGCCCGAGCGCGTCAAGGGACACGACCGGATCG	4023
Db	41151	CCCAGGCGCGCGCACCCCGGACCGCGTCCGCGTCCCGGGGCCCGACCGGAGCTGACCT	41210
Qy	4024	AGTTCGGGCTGCCACGCGGGCCACAGTCTCTCGGCGCCAGCTCCGATCTCGTACGCGTAC	4083
Db	41211	ATCCGAGCTGGACGAGCGCTCCGCGCGCTCCGCGCTGGCTGATTCGGCGCGCGGGTCG	41270
Qy	4084	CAGAGCCCGGTCCGCGCGGAGTCTTGACTGGAAGTGCGCCCGCGGCTGGGTGAGCGCGTTG	4143
Db	41271	CCGCGACACCCCGGTGCGTGTGCTGAGGGCTCCCGCGAGCTGCCGTTGGC-GATC	41329
Qy	4144	CGCCAGGGCGGAAGCGGGGAGTTCGGCGCGGAGCATCATGGGGCGGCGCGTTCGAAA	4203
Db	41330	CTCGCGTGTCTAAGCGCGCGCGGTATCTGCCGATCGATCCGCGCGAGCCGCGCGCGC	41389
Qy	4204	CGGCGCGGCGCGGGTCCAGGGCTCTCTCAAGACGTTCTAGACCCCGCCACCCCGCCCTGC	4263
Db	41390	GCGATCCCGCATCTGTGGCCGACGCGCGCCCGCGCTGTGCTGGCCCCGAGCGTCCACC	41449
Qy	4264	CGCCAGCGCGGACGGCGCGCGCACGCTCGGCGGTTCGGTTCGGGCGGCGGAGG	4318
Db	41450	GCGGACGCTCTGGCCGACAGCTCTCCGGGGCTGTGCTGGCCCCCGCGCTGCGAG	41504

RESULT 10

```

US-11-052-554A-539/C
; Sequence 539, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 539
; LENGTH: 2196
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-539

```

	Query Match	1.6%;	Score 70.8;	DB 8;	Length 2196;
	Best Local Similarity	43.2%;	Pred. No. 0.0078;		
	Matches 677;	Conservative	0;	Mismatches 862;	Indels 29; Gaps 6;
Qy	2755	CAGCGACTACAGGACCTCGAACACCGCTACGCGCTGCTGAGAGACACATCTCTGCCCGC	2814		
Db	1990	CAACGCGCCCTTGCGGCGGTGGCAACCAACGCTGGGCGGCCCGATACGGTTCCGCGGT	1931		
Qy	2815	CATCGCGTTCCCTCCGTGTGGCGCGCTCGATCTCTCCGAGGAGCGCGAGCACGCGCT	2874		
Db	979	CGCACCGTGGAAATCTTCCCGCGGCCCGCGCGCTCCGCGCGCGCCCATCGCCGA	920		
Qy	3827	-----TGCGTAGCGACCCATNGCGGCGCGCTCGGCGCTTCGAATCGACGTCACAG	3877		
Db	919	CCAGGTACCGCGCGCCACCGGCGCGCGCGCTTGTCTGGCGCGCGCGCGCG	860		
Qy	3878	CCTTACTGTGACCGCGTCAGACGGGCGCGAGTGGCGCCCGTTGGACGGCTGGGGCCAGAT	3937		
Db	859	TCTCGGTGTTTTGCCCGCGCGATGCGCGCGGTCCCCCGTGGTCCCGCGCGCGCGACAGCC	800		


```
Qy 219 CGCGGACGATACCCCTTGGCTTCGGCCGAAAGGCTTCGGCGTG--GTCCGCGCCGAGAT 276
Db 29867 CGCGACGACGAAACCGCTGGCGACACCGCGCGGGTTCGGGGGCGAGCGCGCGAGAC 29926
Qy 277 GCGGAACGGC--CTGGCGGGCGCGCGCTCACTTCCCGCACGGTCGGCACGACCTGCGCCT 335
Db 29927 GCGTTCGCGCGAGCGCGCGGAGCGGATGATACCGGGTCGCTGGAGACGATGGCGCT 29986
Qy 336 TGGCCCCCGCGCTGCGCTTCGCGCTCTCAGGAACGGGTTCGGCAACGGGCGCTCGGAT 395
Db 29987 GGAGCGCGGAGCTGGCGCGCGCGAGGACGCTGCTCGCGGTGACGCTGCGCTCGC 30046
Qy 396 ACTTCCGAGATCTGCTCGCGCGCGCGCTTCAGCCCGAGCTCTCGAAGCGAGTAC 455
Db 30047 GCGCGCGCTGGCGCTGGCGCGCGCGCGCTGCGGAGCTGCTCAGGCGCTG--CGCG 30105
Qy 456 GCCCAGCATCTCGAGAGCGCGCGCTGCTGCTGCTGGCGGTGATCGCTTCGGGCGCGG 515
Db 30106 ACGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 30165
Qy 516 ACAGCGGTAGCCCGCTTCTGCTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 575
Db 30166 TCGTTCGGCGCTGACGCGCGCTGCGCGCGCGCTGCGCGCTGCTGCTGCTGCTGCTGCTG 30225
Qy 576 CGATGTCGCGCGATTCAGACGAGCAACCGGGGAGCTGCGCTGACGCGCGCGAGCAGC 635
Db 30226 CCGCGCGCGCTGCGCGCGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 30285
Qy 636 GTTGGACCGATGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 695
Db 30286 GCGCGGACGAGGCGTGGACCGCTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 30345
Qy 696 GGAAGTCCAGTCTAGCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 755
Db 30346 CGCGCGCGAG--TACCCGACCGCGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 30403
Qy 756 TGATGAAGTCTC--CGCGCGCGCTTCCCGACCGCGAGCTGAGACGCTTACACAGTGC 814
Db 30404 GGAAGGCGCTCTTCTGCTGCGCGCGCTGCGCGCGCGCTGCGCGCGCGCTGCGCGCTG 30463
Qy 815 CGGACCGCGCTGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 870
Db 30464 CAGCGCGAGCGCGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTG 30523
Qy 871 ACTGTTCTGACGAGGAACGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 930
Db 30524 GCGCGGACGCGCTGATCCGAGGAACAGCGCTGA--TCAGGGGCGCACCGGGAG 30581
Qy 931 GTCGCGTACTGCGCGCGCGAGACCTCGACGTTGCGCGCGCTGAGCTACCGCTTCGG 990
Db 30582 CTCGGCGGCTGTCGCGGAACCTCTGCTGAGCGCGCGCTGCGCGCGCTGCGCGCTG 30641
Qy 991 CCGCGCGGAGGACTCTGAGGAGACCGTTCGGCGCGAGGCTGGGCGAGCTGACGCGC 1050
Db 30642 ACGTCGCGCGCGGATGAGACGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 30701
Qy 1051 G 1051
Db 30702 G 30702
```

RESULT 13
US-11-075-185-36/c
; Sequence 36, Application US/11075185
; Publication No. US20050266434A1
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185

```
; CURRENT FILING DATE: 2005-03-07  
; PRIOR APPLICATION NUMBER: US 60/551,103  
; PRIOR FILING DATE: 2004-03-08  
; PRIOR APPLICATION NUMBER: US 60/568,290  
; PRIOR FILING DATE: 2004-05-04  
; NUMBER OF SEQ ID NOS: 61  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 36  
; LENGTH: 5679  
; TYPE: DNA  
; ORGANISM: Sorangium cellulosum  
US-11-075-185-36
```

```
Query Match 1.5% Score 66.6; DB 8; Length 5679;  
Best Local Similarity 44.2%; Pred. No. 0.037;  
Matches 694; Conservative 0; Mismatches 849; Indels 28; Gaps 9;  
Qy 1 GTCCGACACCGCGCTCGGGTTCGACGCGCGCGCTACTCGTTCCCGGCCACCGGGCACCG 60  
Db 4791 GTCCAGGAACGTGTTGGCCGCGGGTAGTTGCGCTGCCCGGGTGCACGCGTGCCTGC 4732  
Qy 61 GGTCTGCGCTCGATGACCGACCGCATCGAGAACGTGCGCTGCGGGCTCTCTGATAGTGAGC 120  
Db 4731 CGCGGACGAGAACAGCACGAAAGCGCGAGCGCGCTGCTTCTGAGCTCGTGAGATG 4672  
Qy 121 GGCCCCCGCGCGCACGACATCGACGCTCCATCTCCCTGCTGCGCCCCCG--GAAAGTGAG 179  
Db 4671 AAGGCGCGCTGACACTTTCGCGCGCGACACGCGCGCGCTGCTGCTGCGCGAGAGCTTGGT 4612  
Qy 180 CACCGCACCGCGACGTCGACGCCATGGAGGGCGCACGCTCCCGGACAGCATACCCCTGCGC 239  
Db 4611 CACGACGCTGCTGTCAGCACACCTGCGGTGTCACGACGCGCTCAGGGAGACGCGCGC 4552  
Qy 240 TTGGCGGAAAGGCTTCGCGGTGTCGCGCGCGCATGTCGGAACGCGCTGCGCGGGCGCG-- 297  
Db 4551 CGGGATCGCTTCCACACCGCGCGCGCTGCGCTGCGCGCGCTGCGACGCGCGGAC 4492  
Qy 298 ---GCCCGTCACTTCCCGCAGCGTCGCGACGCTGCGCTTGGCCCGCGCGCTGCGC 354  
Db 4491 GTCGACGCTGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4432  
Qy 355 CTCGCGCGCTTCCAGGAACGCTGCGCGACGCGGCTTCGATATCTTCGAGACATCTGCTC 414  
Db 4431 GTCCATCCCGCGCGCGCGCTCAGCAGAGTTCGCGACGCGCTGCGCGCTGCGCGCTG 4372  
Qy 415 GCGCGCGCTTCGACCGCGCGCTCTCGAAGCAGAGTGACCGCGCGCGCGCTGCGAGAG 474  
Db 4371 CTTCCGACATGCGCGCGCGCGCTTCCCGGTGCGCGCTGATCAGCGCTGTTCTGCTGCGG 4312  
Qy 475 CCGCGCGCTGCTGCTGCTGCGCGGTGATGCTGCGCGCGCGCGACAGCGCTACGCGCGTCC 534  
Db 4311 ATCCAGCGCGCTGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4254  
Qy 535 CTCGTCGCGCGCGCGCTGAGTGCCTGACCGCGCATCTTCGCGGATGTTCGCGGATCGAC 594  
Db 4253 GCGCGCGCGCTGCGCGCTGCGCACGAAAGCTCCGCTCCGCTCCCGCGCGCGCGCGC 4194  
Qy 595 GCAAGCAACCGGGAGCTGTCGCTACGCGCGGACACGCGCTGCGGACCGGATGGCGGG 654  
Db 4193 CGAGCGAA--GAGCGCGCGCTCCAGCGGCTGCGCGCGCGCGCTCCAGCACCGCGCGC 4135  
Qy 655 CGCCGACGACGCTGTTGGACATGAGGTCCTGCGCGCGCGCGAGGCTTCAGTCTTAGGCC 714  
Db 4134 GTCGCGGTACTGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4075  
Qy 715 GGAATCGGTACGCGCGCGCTGCTTCTGCGCTGCGCGCGCT---GATGAAGTGTCTCGC 771  
Db 4074 GTCCAGCGCTGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCGCTGCGCGCT 4015  
Qy 772 GCGCGGTTCCCGCGCGCGCGCGCGCTTACAGGTCGCGGAGCGCGCGCTGCGC 831  
Db 4014 GCGCGCGAGCTGCGCGCGCGCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3955
```


Db 3005 GGGGGCCCTCGTCAGTTCAGACACGACCGCTCGGGCGCGCCCTTCTCCAGC 2946
Qy 4219 TCCAGGGCTCCTCCAGACGTCCTAGACCCGCAACCGCCCTGCGCCAGCGGCGGACG 4278
Db 2945 GCCTCGAGCAGTCCGGCAGCCCGCCAGCGCGAGCGCCCGCCAGCGCCGACGCGCGC 2886
Qy 4279 GCCCGGCGCAGTCCGGCTGTCGGGTCCGGGGCGCGAGG 4318
Db 2885 CCGTCGCGCGCAGCACACGTCGGCGCTCGGCGCGCCAG 2846

RESULT 15

US-11-052-554A-532/c
; Sequence 532, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 532
; LENGTH: 2250
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-532

Query Match 1.5%; Score 66; DB 8; Length 2250;
Best Local Similarity 43.7%; Pred. No. 0.044;
Matches 642; Conservative 0; Mismatches 805; Indels 23; Gaps 7;
Qy 2885 CTGGCACCGACCGGGAAGGACTGACCGCGAAGCGCCCGCACCGGATACCGACCGCGCT 2944
Db 2141 CGTTTACCGAGCCGGAAGCGGCTCCACCGGCGCGCGCTCCCGCGCGCTTCCGCGC 2082
Qy 2945 GCCGAGCGCGGACCGGGGCGCTACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3004
Db 2081 GCCCGCGCGCGCGCGCGCTTACCGCGCTTGCAGATCAGCCCGCGCTTCCCGCGCGCTCCG 2022
Qy 3005 CGTACCGAAGCGTGG- CGGGTCAGAGAATCGTTCCGCGCTGTGGCATCGACGCTACTGGCGC 3063
Db 2021 CAGCGCGCGCGAGTTCGGTTCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 1962
Qy 3064 GTGATCCACCGTGTGTCGGAGGCGCAGAAAGGCCACACGTCGGCGATGTCTCGGGT 3123
Db 1961 AACAGCAGCGCGCGCTCG 1902
Qy 3124 CTGCCGATCGGTTGAACACGAGAGTTGGGGCGAGTCGCGG--GCCGCTCGGGGCT 3181
Db 1901 CCGGCACCTCCGACCCCGCGAGCGGACGACGAGCGCGCGGTTCGCGCGCGCGCGCGCGC 1842
Qy 3182 GCCCGCGCGGTTCATGTCCGTCCTCAGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3241
Db 1841 GCCCGCGCGGTTCAAACCTTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1782
Qy 3242 TCCCGCGTTCGCCAGTTCCTGGCGAGGCGAGCGTGTGAGCGTGTCCACCGCACCTTTGG 3301
Db 1781 GCGACAAGCCCGCTTGC CG -- 1724
Qy 3302 TCATCGGTATCCGATGAGATCGCGGGAACGCGCGCGCGGTTCGCGCAGACGAGATGTTGA 3361
Db 1723 -----CGGATCCG 1670
Qy 3362 TGATCGCGCGCGCGTGC- GCAGTCTGTTTCACTGCTGTGACCGAGAACGAGCGGTGC- 3419

Db 1669 CGGCCCCCGGACGCTTGTGCTCCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1610
Qy 3420 -CCGAGTGTGAAGCGGCGACACGAGTCGTCGAGAGACTCTCTCGGTGACTTTCGTGATCCGTC 3478
Db 1609 GGCACCG 1550
Qy 3479 CCGAGCGCTGAGCGCGCGCGCTTGTTCACAGGATGTGAACTCTCGGGGGCGACTCTCGAACT 3538
Db 1549 CGGCG 1490
Qy 3539 CGCCCATCCCGCGCTCGAAACCGCGCTAGAGCGCGCGCGGTCAACCAAGACGCGCGCGCTT 3598
Db 1489 CGCCCGCGCGGAAACCGCGCTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1430
Qy 3599 CGGCGCGGATGGCCAAACCGCTGTCGCGCGCTGCTCCGATGCTCTGACGCTCTCTCGG 3658
Db 1429 CAAACAGCAAGGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTC 1370
Qy 3659 CCGCGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3718
Db 1369 CGGCG 1310
Qy 3719 GGGCGATACCGCGTCCGATGCGCGGTTCCTCCCGGTACACAGGGCG-----GTCTTGC 3772
Db 1309 CTCGCGCGCGCGCGCGCTGACGCGCGACCGCGCGGTCTCTCCCGCGCGCGCGCGCGCG 1250
Qy 3773 CTTACGGGTCTTCATACCTGCTCCATGTGACGATATCAGCCCCCGCGCGGTGCGT 3832
Db 1249 CGCGTGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTCGCGATCAGCG 1190
Qy 3833 AGCGACCATGCGCGCGCTCGCGCGCTTGAATCGACGCTCAGACGCTTACCTGTGACCGCG 3892
Db 1189 GCGCGCGCTCAGCGCTTGAATGCGCGGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCG 1130
Qy 3893 GTCAGACGGCGCGCGGTGCGCGGTTCGACGCGTGGGGCGCAGATC---GGGCGCGCGCG 3949
Db 1129 TGTGACGCGGTGCTGCG 1070
Qy 3950 ACGGGGAACCGCGCGCGTCAAGGGTCAAGGGTTCGCGGGACCGCGCGCGCGCGCGCGCG 4009
Db 1069 CGATGCG 1010
Qy 4010 CACCGACCGGATCGAGTTCGCGCGTGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4069
Db 1009 CGCGTTCGCGATCAGCG 950
Qy 4070 ACTGTAAGCGTACAGAGCGCGGTTCGCGCGCGAGTCTGAGTGTGACGTGGCGCGCGGT 4129
Db 949 CCGCGGTGGCAACCGAGCTGCGCGCGCTGCCACCGTTGCGCGCGTTCGCGATCAGCGCGG 890
Qy 4130 GGTGAGCGGTTCGCGCGCGGCGGAGGCGCGGAGTTCGCGCGCGCGCGCGCGCGCGCGCG 4189
Db 889 CCG 830
Qy 4190 GGGCGCGTTCGAAACCG 4249
Db 829 CCGACCG 770
Qy 4250 CAAACCGCGCGTTCG 4309
Db 769 CCGCGCGGAGTTACCGGTGTACCGAAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 710
Qy 4310 CGGCGCGGCGAGCGACCGGTAGAGCGCG 4339
Db 709 CACCGGAGCGCGAAGCAGCG 680

Search completed: January 19, 2006, 04:10:04
Job time : 628 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2006, 16:18:41 ; Search time 3182 Seconds
(without alignments)
11294.380 Million cell updates/sec

Title: US-10-017-471B-19
Perfect score: 4346
Sequence: 1 gtcagcagcggtcggtggtt.....ggtagagcgcggtacc 4346

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA_Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	4346	100.0	4346	US-10-017-471A-19
2	401	9.2	401	US-10-017-471A-12
3	401	9.2	401	US-10-017-471A-13
4	377.2	8.7	9025608	US-10-156-761-1
5	371.2	8.5	840	US-10-156-761-3512
6	316.2	7.3	19016	US-10-469-442-1
7	274.2	6.3	88400	US-10-844-716-1
8	255.4	5.9	7600	US-10-732-859-1
9	253.6	5.8	9025608	US-10-156-761-1
10	250	5.8	2370	US-10-156-761-1965
11	236.8	5.4	906	US-10-156-761-924
12	229	5.3	699	US-10-049-710A-1
13	213	4.9	741	US-10-369-493-43039
14	210.6	4.8	762	US-10-369-493-28527
15	210.6	4.8	762	US-10-369-493-31286
16	206.6	4.8	750	US-10-369-493-35359
17	206.6	4.8	750	US-10-369-493-37943
18	206.6	4.8	750	US-10-369-493-38141
19	206.6	4.8	750	US-10-369-493-38559
20	185.4	4.3	732	US-10-369-493-35573
21	176.6	4.1	777	US-10-369-493-35593
22	174.6	4.0	762	US-10-369-493-37608
23	172.8	4.0	765	US-10-369-493-39254

c 24	172.8	4.0	765	6	US-10-369-493-39627	Sequence 39627, A
c 25	172.6	4.0	747	6	US-10-369-493-35639	Sequence 35639, A
c 26	171.4	3.9	747	6	US-10-369-493-35942	Sequence 35942, A
c 27	170	3.9	702	6	US-10-156-761-3692	Sequence 3692, Ap
c 28	169.2	3.9	753	6	US-10-369-493-40200	Sequence 40200, A
c 29	167.6	3.9	732	6	US-10-369-493-39993	Sequence 39993, A
c 30	164	3.8	726	6	US-10-369-493-30920	Sequence 30920, A
c 31	162.8	3.7	720	6	US-10-369-493-28161	Sequence 28161, A
c 32	162	3.7	765	6	US-10-369-493-24326	Sequence 24326, A
c 33	160.6	3.7	1029	7	US-10-437-963-98530	Sequence 98530, A
c 34	158	3.6	742	6	US-10-369-493-44176	Sequence 44176, A
c 35	154	3.5	699	6	US-10-369-493-32280	Sequence 32280, A
c 36	151.2	3.5	852	7	US-10-437-963-64696	Sequence 64696, A
c 37	151	3.5	663	6	US-10-156-761-2353	Sequence 2353, Ap
c 38	150.2	3.5	981	7	US-10-425-114-15642	Sequence 15642, A
c 39	150.2	3.5	1138	7	US-10-425-114-15867	Sequence 15867, A
c 40	150.2	3.5	1385	7	US-10-425-114-15463	Sequence 15463, A
c 41	150.2	3.5	1448	8	US-10-425-115-9925	Sequence 9925, Ap
c 42	148.4	3.4	819	6	US-10-156-761-4765	Sequence 4765, Ap
c 43	148.2	3.4	735	6	US-10-369-493-30930	Sequence 30930, A
c 44	148.2	3.4	780	6	US-10-369-493-28171	Sequence 28171, A
c 45	147	3.4	979	7	US-10-425-114-34802	Sequence 34802, A

ALIGNMENTS

RESULT 1
US-10-017-471A-19
; Sequence 19, Application US/10017471A
; Publication No. US200301246441
; GENERAL INFORMATION:
; APPLICANT: Takano, Eriko
; APPLICANT: Bibb, Mervyn
; TITLE OF INVENTION: Antibiotic Production
; FILE REFERENCE: 0380-P02329US1
; CURRENT APPLICATION NUMBER: US/10/017, 471A
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/242,561
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 4346
; TYPE: DNA
; ORGANISM: Streptomyces coelicolor
US-10-017-471A-19

Query Match	100.0%	Score 4346;	DB 6;	Length 4346;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 4346;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GTGAGGAGCGGCTCGGGTTCCACCGGAGCGGTACTCGTTCCCGGCCACCGGGCACCG	60	
DB	1	GTGAGGAGCGGCTCGGGTTCCACCGGAGCGGTACTCGTTCCCGGCCACCGGGCACCG	60	
QY	61	GGTCTGCGCTCGATGACCGACCGCATCGAGGAGCGTGGCTGGCGGCTCTCTGATAGTGAGC	120	
DB	61	GGTCTGCGCTCGATGACCGACCGCATCGAGGAGCGTGGCTGGCGGCTCTCTGATAGTGAGC	120	
QY	121	GGCCCCCGCGGGGACCGACATCGACGTCATCTCCCATCTGCGCCCCCGGAAAGTGAGC	180	
DB	121	GGCCCCCGCGGGGACCGACATCGACGTCATCTCCCATCTGCGCCCCCGGAAAGTGAGC	180	
QY	181	ACCGCACCGGAGCGTGGAGGCGGCGGATCGGCGGAGCGGATCACCCTGGCT	240	
DB	181	ACCGCACCGGAGCGTGGAGGCGGCGGATCGGCGGAGCGGATCACCCTGGCT	240	
QY	241	TCGGCGGAAGGCTTCGCGTGGTCCGCCGCCACGATGCGGAACCGGCTGGCGGGCGGGCC	300	
DB	241	TCGGCGGAAGGCTTCGCGTGGTCCGCCGCCACGATGCGGAACCGGCTGGCGGGCGGGCC	300	
QY	301	CGTCACTTCCCGACGCGTGGCACGACCTGCGCCTTTGGCCCCCGCGCTGCGCTGCGC	360	

Qy	2521	CTGTCACCGCTTGC	CGCACGAAACGTCGTGGCCCGGGCCGGCGTGGCGCTCTCCATGACCA	2580
Db	2521	CTGTCA	CGCGTTGCGCACGAAACGTCGTGGCCCGGGCCGGCGTGGCGCTCTCCATGACCA	2580
Qy	2581	GCAGGCGCACGGTCT	CGATCGCCCGAGAACCTTCCGTCGTGGCAGAGACATCTCTGAA	2640
Db	2581	GCAGGCGCACGGTCT	CGATCGCCCGAGAACCTTCCGTCGTGGCAGAGACATCTCTGAA	2640
Qy	2641	GCTGCTGAAACGAGCC	CAAGGAGAACGCTGAGTTGCTGCCCCATGTGGTCACCAACCGACTC	2700
Db	2641	GCTGCTGAAACGAGCC	CAAGGAGAACGCTGAGTTGCTGCCCCATGTGGTCACCAACCGACTC	2700
Qy	2701	GGCCGATCTCTTA	CGTGGGCGACGTTTCGCGGGGATACAGTCTGTGTCCAGACGCTCAGCGA	2760
Db	2701	GGCCGATCTCTTA	CGTGGGCGACGTTTCGCGGGGATACAGTCTGTGTCCAGACGCTCAGCGA	2760
Qy	2761	CTACACGAGACCT	CGAACACCGCTACAGCGCTGCTGCAGAGACATCTCTGCCCGCATCGC	2820
Db	2761	CTACACGAGACCT	CGAACACCGCTACAGCGCTGCTGCAGAGACATCTCTGCCCGCATCGC	2820
Qy	2821	GGTTCCCTTCGCTG	CTGGCCGCTCGATCTCTCCGAGGAGCGGGAGACAGCTCTCGCGC	2880
Db	2821	GGTTCCCTTCGCTG	CTGGCCGCTCGATCTCTCCGAGGAGCGGGAGACAGCTCTCGCGC	2880
Qy	2881	CGAACTGGCACCG	ACCGGGAAGACTGACCGCCGGAAGCGCCGACCGGATACCGACCCG	2940
Db	2881	CGAACTGGCACCG	ACCGGGAAGACTGACCGCCGGAAGCGCCGACCGGATACCGACCCG	2940
Qy	2941	CCGTGCCGAGCG	CGCCGACCGGGCCCGCTTACGGGCCCGGGCGCGGCCCGTGGTCTGC	3000
Db	2941	CCGTGCCGAGCG	CGCCGACCGGGCCCGCTTACGGGGCCCGCTTACGGGGCCCGCGCGGCCCGTGGTCTGC	3000
Qy	3001	CCTGCGTACCGA	AGCGTGGCGGTACAGAAATCGTTCCGCTGTGGCATCGAGTACTCG	3060
Db	3001	CCTGCGTACCGA	AGCGTGGCGGTACAGAAATCGTTCCGCTGTGGCATCGAGTACTCG	3060
Qy	3061	CCGCTGATCCAC	CGCTGAGTCTCGGAGGCGCAGAAAGCCACACGTCGGCGCATGTCTCG	3120
Db	3061	CCGCTGATCCAC	CGCTGAGTCTCGGAGGCGCAGAAAGCCACACGTCGGCGCATGTCTCG	3120
Qy	3121	GGTCTGCCGATG	CGGTTGAACAACGAGTGTGGCGCCAGTGGCGCGCGCTCGGGGTC	3180
Db	3121	GGTCTGCCGATG	CGGTTGAACAACGAGTGTGGCGCCAGTGGCGCGCGCTCGGGGTC	3180
Qy	3181	TGCCCGCGCGTGT	GTTCATGTCTCCACGAAACCCGGCGCCACCGGGTTGACCGTG	3240
Db	3181	TGCCCGCGCGTGT	GTTCATGTCTCCACGAAACCCGGCGCCACCGGGTTGACCGTG	3240
Qy	3241	ATCCCCCGTTCC	CCAGTTGCTGGCCAGGGCGAGCGTGACGCTGTCCACCGACCCCTTG	3300
Db	3241	ATCCCCCGTTCC	CCAGTTGCTGGCCAGGGCGAGCGTGACGCTGTCCACCGACCCCTTG	3300
Qy	3301	GTCAATCGCGTA	TCGATCGGGAACGCGCGCGGGTTCGCGGAGACGAGATGTG	3360
Db	3301	GTCAATCGCGTA	TCGATCGGGAACGCGCGCGGGTTCGCGGAGACGAGATGTG	3360
Qy	3361	ATGATTCGCGCC	CGCGTTCGCGAGTCTGTTGAGTCCGTGGACACGAGCGGTGCC	3420
Db	3361	ATGATTCGCGCC	CGCGTTCGCGAGTCTGTTGAGTCCGTGGACACGAGCGGTGCC	3420
Qy	3421	CGGACGTTGAC	GGCGACAGTCCGCTCGGACCTCTCCGTGATCCGTC	3480
Db	3421	CGGACGTTGAC	GGCGACAGTCCGCTCGGACCTCTCCGTGATCCGTC	3480
Qy	3481	GAGCGGCTGAC	CGCCCGCTTCTTACACGAGATGTCGAACTCGGGCGGCACTCGAATCG	3540
Db	3481	GAGCGGCTGAC	CGCCCGCTTCTTACACGAGATGTCGAACTCGGGCGGCACTCGAATCG	3540
Qy	3541	CCCATCCCGCGT	CGAACCGCGGTAGAGCGCGCGCGTCAACGACGAGCTTCG	3600
Db	3541	CCCATCCCGCGT	CGAACCGCGGTAGAGCGCGCGCGTCAACGACGAGCTTCG	3600

Qy	3601	GC	CGGATG	GC	CAACGC	CTGT	CCGC	CTGCT	CTGGAT	TGTC	CTCG	AGCGT	CT	CTCG	GC	3660
Db	3601	GC	CGGATG	GC	CAACGC	CTGT	CCGC	CTGCT	CTGGAT	TGTC	CTCG	AGCGT	CT	CTCG	GC	3660
Qy	3661	GC	CGCT	CT	CGT	TG	CGT	AGT	TG	CACTG	CCAC	GAGCG	CCCC	CGTCCG	CGGCG	3720
Db	3661	GC	CGCT	CT	CGT	TG	CGT	AGT	TG	CACTG	CCAC	GAGCG	CCCC	CGTCCG	CGGCG	3720
Qy	3721	GC	GATAC	CG	CGT	CC	GATG	CC	CGCT	T	CCCC	CGGT	CA	CA	AGGCG	3780
Db	3721	GC	GATAC	CG	CGT	CC	GATG	CC	CGCT	T	CCCC	CGGT	CA	CA	AGGCG	3780
Qy	3781	GGT	CTT	CCA	TAC	CT	CTG	T	CCATG	TG	CA	CG	CA	GTAT	CAG	3840
Db	3781	GGT	CTT	CCA	TAC	CT	CTG	T	CCATG	TG	CA	CG	CA	GTAT	CAG	3840
Qy	3841	AT	GGCG	CG	CGT	CG	CGT	T	CGAAT	CG	AC	CGT	CA	CA	CG	3900
Db	3841	AT	GGCG	CG	CGT	CG	CGT	T	CGAAT	CG	AC	CGT	CA	CA	CG	3900
Qy	3901	GGG	CCG	GAGT	TG	CG	CGT	T	TG	GA	CG	CGT	TG	GG	CG	3960
Db	3901	GGG	CCG	GAGT	TG	CG	CGT	T	TG	GA	CG	CGT	TG	GG	CG	3960
Qy	3961	GC	CG	CGT	CA	GG	GGT	CG	CG	GA	CC	CG	CA	CG	CG	4020
Db	3961	GC	CG	CGT	CA	GG	GGT	CG	CG	GA	CC	CG	CA	CG	CG	4020
Qy	4021	T	CG	AGT	TC	GG	CGT	CG	CG	CA	CC	AGT	CT	CT	CG	4080
Db	4021	T	CG	AGT	TC	GG	CGT	CG	CG	CA	CC	AGT	CT	CT	CG	4080
Qy	4081	T	AC	CAG	AG	CC	CG	CT	CG	CG	CA	CGT	T	CG	CG	4140
Db	4081	T	AC	CAG	AG	CC	CG	CT	CG	CG	CA	CGT	T	CG	CG	4140
Qy	4141	T	TG	CG	CA	GG	CG	GA	AG	CG	GG	AGT	CG	CG	CG	4200
Db	4141	T	TG	CG	CA	GG	CG	GA	AG	CG	GG	AGT	CG	CG	CG	4200
Qy	4201	A	A	AG	CG	CG	CG	CG	CGT	CC	A	GG	CGT	CT	C	4260
Db	4201	A	A	AG	CG	CG	CG	CG	CGT	CC	A	GG	CGT	CT	C	4260
Qy	4261	TG	CG	CA	GG	CG	CG	CG	CGT	CT	CG	CG	CGT	CT	CG	4320
Db	4261	TG	CG	CA	GG	CG	CG	CG	CGT	CT	CG	CG	CGT	CT	CG	4320
Qy	4321	AG	CG	AC	CG	T	A	G	CG	CG	CG	GGT	T	A	C	4346
Db	4321	AG	CG	AC	CG	T	A	G	CG	CG	CG	GGT	T	A	C	4346

RESULT 2

```

US-10-017-471A-12
; Sequence 12, Application US/10017471A
; Publication No. US20030124644A1
; GENERAL INFORMATION:
; APPLICANT: Takano, Eziko
; APPLICANT: Bibb, Mervyn
; TITLE OF INVENTION: Antibiotic Production
; FILE REFERENCE: 0380-P023290S1
; CURRENT APPLICATION NUMBER: US/10/017,471A
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 6/242,561
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Streptomyces coelicolor
US-10-017-471A-12

```


Query Match 9.2%; Score 401; DB 6; Length 401;
Best Local Similarity 100.0%; Pred. No. 8.7e-86;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1973 GGGCAGGACGGCGGTGACCGAGAACCGGTTCACCGCCCTTCGGTATCCAGCTGACCGGAA 2032
Db 1 GGGCAGGACGGCGGTGACCGAGAACCGGTTCACCGCCCTTCGGTATCCAGCTGACCGGAA 60

Qy 2033 CGCGTCTCTGCACCTTGGTCCGGTGGACAAAGCGCCATCGGAACCGCAATGCGGTTTGTTC 2092
Db 61 CGCGTCTCTGCACCTTGGTCCGGTGGACAAAGCGCCATCGGAACCGCAATGCGGTTTGTTC 120

Qy 2093 GATCGAGTTGGATCGGACGAGCAAGATTGATCAAACTACTGCTTCGGGCATGGTCCCCC 2152
Db 121 GATCGAGTTGGATCGGACGAGCAAGATTGATCAAACTACTGCTTCGGGCATGGTCCCCC 180

Qy 2153 CCAGGAATCATGTGATCGCGAGCTGTTCTGTATGCGGAACGTTAAGATACAGACTGAGC 2212
Db 181 CCAGGAATCATGTGATCGCGAGCTGTTCTGTATGCGGAACGTTAAGATACAGACTGAGC 240

Qy 2213 GGTTTTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGATGCGCCAAAGCAG 2272
Db 241 GGTTTTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGATGCGCCAAAGCAG 300

Qy 2273 GACCGGGCGATCCGACCGCGGAGACGATCTTGAGCGCGCGGCGAGGTTCTCGAGAG 2332
Db 301 GACCGGGCGATCCGACCGCGGAGACGATCTTGAGCGCGCGGCGAGGTTCTCGAGAG 360

Qy 2333 CAGGGCTACCAAGTGCACACGATCAGCGAGATCTCAAGGT 2373
Db 361 CAGGGCTACCAAGTGCACACGATCAGCGAGATCTCAAGGT 401

RESULT 3
US-10-017-471A-13/c
; Sequence 13, Application US/10017471A
; Publication No. US20030124644A1
; GENERAL INFORMATION:
; APPLICANT: Bibb, Mervyn
; APPLICANT: Takano, Eriko
; TITLE OF INVENTION: Antibiotic Production
; FILE REFERENCE: 0380-P02329US1
; CURRENT APPLICATION NUMBER: US/10/017,471A
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/242,561
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Streptomyces coelicolor
US-10-017-471A-13

Query Match 9.2%; Score 401; DB 6; Length 401;
Best Local Similarity 100.0%; Pred. No. 8.7e-86;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1973 GGGCAGGACGGCGGTGACCGAGAACCGGTTCACCGCCCTTCGGTATCCAGCTGACCGGAA 2032
Db 401 GGGCAGGACGGCGGTGACCGAGAACCGGTTCACCGCCCTTCGGTATCCAGCTGACCGGAA 342

Qy 2033 CGCGTCTCTGCACCTTGGTCCGGTGGACAAAGCGCCATCGGAACCGCAATGCGGTTTGTTC 2092
Db 341 CGCGTCTCTGCACCTTGGTCCGGTGGACAAAGCGCCATCGGAACCGCAATGCGGTTTGTTC 282

Qy 2093 GATCGAGTTGGATCGGACGAGCAAGATTGATCAAACTACTGCTTCGGGCATGGTCCCCC 2152
Db 281 GATCGAGTTGGATCGGACGAGCAAGATTGATCAAACTACTGCTTCGGGCATGGTCCCCC 222

Qy 2153 CCAGGAATCATGTGATCGCGAGCTGTTCTGTATGCGGAACGTTAAGATACAGACTGAGC 2212

Db 221 CCAGGAATCATGTGATCGCGAGCTGTTCTGTATGCGGAACGTTAAGATACAGACTGAGC 162

Qy 2213 GGTTTTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGAGCATGGCCCAAGCAG 2272

Db 161 GGTTTTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGAGCATGGCCCAAGCAG 102

Qy 2273 GACCGGGCGGATCCGACGCGGAGACGATCTTGAGCCCGCGGCGAGGTTCTTCGAGAAG 2332

Db 101 GACCGGGCGGATCCGACGCGGAGACGATCTTGAGCCCGCGGCGAGGTTCTTCGAGAAG 42

Qy 2333 CAGGGCTACCAAGTGCACGATCAGCGAGATCTCAAGGT 2373

Db 41 CAGGGCTACCAAGTGCACGATCAGCGAGATCTCAAGGT 1

RESULT 4
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 8.7%; Score 377.2; DB 6; Length 9025608;
Best Local Similarity 64.1%; Pred. No. 2.8e-80;
Matches 568; Conservative 0; Mismatches 318; Indels 0; Gaps 0;

Qy 227 GATCACCCCTGGCTTCGGCCGAAAGGCTTCGGGTGTCGGCGGCCAGATGCGGAACGCGC 286

Db 4362444 GAGCATGCCGCATTCAGGCGAAGCGGTTCTGTGTGTCGTCGCCACTGTGCTAGGAG 4362385

Qy 287 CTGGCGGGCGGCCCTCAGCTTCCTCGCACGCTTCGCGACGACTGCGCTTCGGCCCCCGCC 346

Db 4362384 CGTGTGGGTGGCCCGTGACCCGGCGGAGACGTTGGGTGGTCTGTCGTCTTCGGGCCACG 4362325

Qy 347 CGCTGCCGCTCGGGCGCTCTCCAGGAAACGCTCGGCGACGCGGCCCTCGGATATTCGAGAGC 406

Db 4362324 AGCTGGCGGTTGTCGCTCAGGACGAGTGTGTCGTCACCACTTTGGGCGGTAGCGCGGCC 4362265

Qy 407 ATCTGCTCGCGCGCGCTTCAGCCCCAGCTCTCTCGAACAACGAGTGTGACCGCCCGGACACC 466

Db 4362264 CACCGGGCGGAGCGCTGCTGTGGTGTCTCAGTTTCGCGGAGACACGCGGGCGCGGAGGACC 4362205

Qy 467 TCGGAGACCGCGCGCTCTGCTGCTGCGGTGTATCGCTCGGGCCCGGACAGCGCGCTAC 526

Db 4362204 GCCGCAGGTGCGCGCTGTGTGCTCCCGGAGAGCCGTTTCGGGGCCGTTGAGGGGATAC 4362145

Qy 527 GCCCGTCCCTCGTGGCCCGGGCGGGGTCTAGTCCCTTGACCCGCACTTCGCGGATGTTCGGCC 586

Db 4362144 ACCCGGCCCTCGTGGCCGGGTGAGGTCTAGCGCGGTTCACCGGACCTCGGCGATGTCTACGG 4362085


```
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      5.8%; Score 253.6; DB 6; Length 9025608;
Best Local Similarity 80.1%; Pred. No. 8.4e-51;
Matches 298; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 3972 GGGTCAGGGGTCGCGGACCGCCCGCCAGCGCGGTTCAGGGCAGCCGATCGAGGTCGGG 4031
Db 2413017 GGTTCACGCTTCGTCCGTGCGCGAGACCCGTCAGGGCGCGCCGAGGTCAGGTCGG 2413076

QY 4032 CGTGCCACCGGCCACCAAGTCCTCGCGGCCAGCTCGAGTCGTCGATCGGTACGAGAGCCC 4091
Db 2413077 GGTGCCCCGGGGCCACCAAGTCGTGTCGTCGCGTTCCGATTCTGATGGCGTTACAGAGGCC 2413136

QY 4092 GGTCCGCGCGAGTCGACCTGACGTGCGCGCGCGGGTGGGTGAGGCGGTTGCGCCAGGG 4151
Db 2413137 GTCGTGGCGGAGGCGGAGCTGGGCAATGCCCCTTCGATGGGTGAGGCGGTTGCGCCAGG 2413196

QY 4152 GCGGAAGCGCGGAGGTCGCGGCGGAGCATCATGGGCGCGCGCGGTGCGAAACGCGCCGCG 4211
Db 2413197 CGGNACCGCGGAGGTCGGCAGCGAGGAGCGGGCGGCGCGGTGCGAAACGTCGCGC 2413256

QY 4212 CGGCGGGTCCAGGCGCTCTCCAGGACGTCTAGACCCGCCAACCGCCCTGCGCGCAGGC 4271
Db 2413257 CGGCGGATCCAGGGTTCTTCGAGGACGCGCGAGCCCTTCGAGTCGCGCCCTGTGCGCAGGC 2413316

QY 4272 GCGGACGCGCGCGCCACACTCGCCGTCGTGCGGTCCGCGCGCGCGGCGAGGCGCGGTA 4331
Db 2413317 GCGGACCGCGCGCGGCGCAAGTCCCCGGTGGTCGTCGCGCGCGCGGTGGCCAGCGCGTA 2413376

QY 4332 GAGCGCGCGGGT 4343
Db 2413377 CAGCGCGCGGTGT 2413388

RESULT 10
US-10-156-761-1965/c
; Sequence 1965, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(906)
US-10-156-761-924
```

```
; SEQ ID NO 1965
; LENGTH: 2370
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2370)
US-10-156-761-1965

Query Match      5.8%; Score 250; DB 6; Length 2370;
Best Local Similarity 80.7%; Pred. No. 8.4e-50;
Matches 292; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 3982 TCGCCGGGACCGCCAGCGCGGTTCAGGGCAGCCGATCGAGTCCGGGTCCGGGTCCCGGC 4041
Db 2366 TCGTCCGGTGGCGGAGACCGCGTCAGGGCGCGCCGAGTCCGGGTCCGGGTCCCGGC 2307

QY 4042 GGCACACAGTCTCTCGCGGCCCGAGCTCCGACTCTGACGCTACCGAGAGCCCGGTCCGGCG 4101
Db 2306 GGCACACAGTCTCTCGGTTCGGTTCGATTCGTAGGCGTACCGAGGCGCGTTCGTGGCG 2247

QY 4102 AGTCTGAGCTGGACGTGCGCGCGCGGTGGGTGAGGCGGTTGCCCGCAGGGCGGAGGCG 4161
Db 2246 AGCGGAGCTGGGCATGCGCCCTTCGATGGGTGAGGCGGTTGCCCGCAGGGCGGAGGCG 2187

QY 4162 GGGAGGTCGGCGGCGAGCATCATGGGCGCGCGGTCGAAACGCGCGCGCGCGGTCC 4221
Db 2186 GGGAGGTCGGCAGCAGGAGAGCGGGCGCGCGGTGAAACGCTCCGGCGCGCGATCC 2127

QY 4222 CAGGGCTCCTCCAGGAGTCTTAGACCCCGCCCAACCCGCCCTCGCCCGCAGCGCGGAC 4281
Db 2126 CAGGGTTCCTCGAGGACGGCGAGCCCTTCGAGTCCGCCCTGTGCGCAGGCGCGACGCC 2067

QY 4282 CGCGCCAGTCCCGCGGTGTCGCTCCGCGCGCGCGAGGCGGAGCGAGCGGTAGAGCGCGCG 4341
Db 2066 CGGCGCAAGTCCCGCGGTGTCGCTCCGCGCGCGGTGGCCAGCGAGCGCGGTACAGCGCG 2007

QY 4342 GT 4343
Db 2006 GT 2005

RESULT 11
US-10-156-761-924/c
; Sequence 924, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 924
; LENGTH: 906
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(906)
US-10-156-761-924

Query Match      5.4%; Score 236.8; DB 6; Length 906;
```

[illegible]

RESULT 12
US-10-049-710A-1
; Sequence 1, Application US/10049710A
; Publication No. US20030136648A1
; GENERAL INFORMATION:
; APPLICANT: Shinmyo, Atsuhiko
; APPLICANT: Kato, Kou
; APPLICANT: Yamada, Yasuhiro
; APPLICANT: Nihira, Takuya
; APPLICANT: Shindo, Takuya
; TITLE OF INVENTION: METHOD FOR INDUCTION OF GENE EXPRESSION IN PLANT AND PLANT TREATMENT
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 5405/18

```

; CURRENT APPLICATION NUMBER: US/10/049,710A
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: PCT/JP01/05096
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: JP 2000-180466
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Streptomyces virginiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(699)
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS: Okamoto, S., Nakamura, K., Nihira, T. and Yamada, Y.
; TITLE: Virginiae butanolide binding protein from Streptomyces virginiae.
; TITLE: Evidence that VbrA is not the virginiae butanolide binding protein and re-
; TITLE: Identification of the true binding protein
; JOURNAL: Journal of Biological Chemistry
; VOLUME: 270
; ISSUE: 20
; PAGES: 12319-12326
; DATE: 1995-05-19
; DATABASE ACCESSION NUMBER: D32251
; DATABASE ENTRY DATE: 1994-07-19
; RELEVANT RESIDUES: (1)..(699)
; PUBLICATION INFORMATION:
; AUTHORS: Okamoto, S., Nakamura, K., Nihira, T. and Yamada, Y.
; TITLE: Virginiae butanolide binding protein from Streptomyces virginiae.
; TITLE: Evidence that VbrA is not the virginiae butanolide binding protein and re-
; TITLE: Identification of the true binding protein
; JOURNAL: Journal of Biological Chemistry
; VOLUME: 270
; ISSUE: 20
; PAGES: 12319-12326
; DATE: 1995-05-19
; DATABASE ACCESSION NUMBER: D32251
; DATABASE ENTRY DATE: 1994-07-19
; US-10-049-710A-1

Query Match 5.3%; Score 229; DB 6; Length 699;
Best Local Similarity 60.8%; Pred.No.9e-45;
Matches 373; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

Qy 2270 CAGGACGGGCGGATCCGACGGGGCAGAGCATCTCGAGCGCCGGCGGAGGTCCTTCGAG 2329
Db 34 CAGGAACGGGCGCGTCCGACCGGGCAGGCGATCGTCGGGCGAGCCGCTCGGTCCTTCGAC 93

Qy 2330 AAGCAGGGGTACCAAGCTGCCACGATCACGGAGATCCTCAAGTGGCCGGGGTGACCAAG 2389
Db 94 GAGTACGGGTTTCAGGCGCCGCCACAGTGGCAGAGATCCTCTCGGGGGGCTTCGGTCAACAAG 153

Qy 2390 GGAGCCCTCTACTTTCCAGTTCCAGTCCAGAGGAAGAACTGGCGCTGGCGGCTCTTCGACGCC 2449
Db 154 GGGCGGATGTACTTCCACTTCGTTTCAGAGGAGAGCTGGCCCGCGGCTGTGGCCGAG 213

Qy 2450 CAGGAACACCAACACAGGCGCGTTCGGAGCAACCCCTCCGGTGCAAGAACTCATCGACATG 2509
Db 214 CAGACCTCAGCATGTCGGGTCGGGAATCCGGCTCCAAAGGCGCAGGAATGGGTAGACCTC 273

Qy 2510 GGATGTTGTTGTGTACCGCTTCGGCAGCAAGCTGTGGCCCGCGGCGGTCGCGCTC 2569
Db 274 ACCATGTGTGTCCCCACGCGCATGCTGCACGATTCGATTCCTGCGGGCGGCGACGCGGCTC 333

Qy 2570 TCCATGGACAGCAGGCGCAGCGTCTCGATTCGCGGAGGACCTTTCGTCGCTGGCACGAG 2629
Db 334 GCACCTGGACAGGGGGCGGTGGACTTCTTCGACGCCAACCCGTTTCGCGAGTGGGGCGAC 393

Qy 2630 ACACCTCTGAAGCTGTGTGAACCAAGGGCCAGGAGAACGGTGAGTTGCTGCCCCCATGTGGTC 2689

```



```
Db 524 ACGAGCGTATCGACCGCGCCCTTGGTTCGGCGCATAGGCGCGAGCGTGCACCGCGGG 465
QY 3337 CCGGTCGGCGCAGACGAGATGTTGATGATCCCGCGCGCTCGCGCAGTCTGTTTCAGTCGG 3396
Db 464 CCGCGCGCGAGCGACGAGAGCAGCACCGCTGCTGCTTGCATCATCACCGCGCAGCAAC 405
QY 3397 TCGTGGACAGAGAACAGCGGTCGCCGCGAGCTTTGAGACGGCGACAGTCGGTCGGAAGACCTCC 3456
Db 404 TCGTGCACGAGGAAGTAGGGTTCGGCGCAGCTTTCAGCGCGCAACAGATCGTCGAAAGTCTTCC 345
QY 3457 TCGGTGACTTCGTCGATCCGTCGCGAGCGCTGACGCGCGCGTGTTCACCAAGGATGTCG 3516
Db 344 ACGGTCGTGTCCTCGATGTCGGCGCTTCGAAATCCCGGCGTTCGCCACCAAGATATCC 285
QY 3517 AACTCGGGCGGCACATCCGAACTCGCCATCCCGCGGTTCGAAAGCGCGCTAGAGCGCGGCC 3576
Db 284 AGCGCGTG-----GCCGATCACCGCGCGCACCTGCGCGCGCAGCTTGTGC 240
QY 3577 GCGTCACCAACAGCGCGAGTTCCGCGCGGATGCGCAAGCCTGTCGCGCGCTGCTCGCG 3636
Db 239 GGGCGGTCGGCTCGCGCAGATTCGCGCGCATTTCTGCGCATGTCGCGCGCGCGCGCA 180
QY 3637 ATGGTCTGACAGGTCTCTCGCGCGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3696
Db 179 ATTTCCGCGACCAACCGCATCCGCTTCTCGCCACTGCTGTAGTGAACCAAGCACCTGT 120
QY 3697 GCCCGCTCGCGCGCAGCGCGGCGATACCGCGTCCGATGCCCGCTTCCCGCGTCC 3756
Db 119 GCGCGGCTCGCGCGAGCGCGAGTGGCTGGCGCGCGCGATGCGCGGCGCGCGTG 60
QY 3757 ACCAGGCGGTCTTGCC 3773
Db 59 ACGAGCGCGTCTTGCC 43

RESULT 15
US-10-369-493-31286/c
; Sequence 31286, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31286:
; LENGTH: 762
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
US-10-369-493-31286
```

```
Query Match 4.8%; Score 210.6; DB 6; Length 762;
Best Local Similarity 57.7%; Pred. No. 2.2e-40;
Matches 425; Conservative 0; Mismatches 294; Indels 18; Gaps 2;

QY 3037 CCGCTGTGGCATCGACGTACTGGCGGTGATCCACGGTGAAGTCGTCGCGAGCGCCAGAAAG 3096
Db 761 CCACCGTCGACATGCAAGGTCTCGCGGTAATCCAGGTCGCGCGTCCGAGGCGAGAAAG 702
QY 3097 GCCACCACTGCGCGATGTCGTCGGTCTGCCGATCGGTTGAACACGAGATTGGCGGCC 3156
Db 701 GCCACGGCGCGCATATCATCTCGGTCGGCCACGCGCTTGAGCGCCTGTCATGCCGAGC 642
QY 3157 AGTCCGCGCGCGCTCGGGGTCTCGCGCGGTCTCATGTCGCTTCCACGAAA 3216
```

Search completed: January 19, 2006, 03:59:39
Job time : 3191 secs

this page blank (uspto)

this page blank (uspto)

this page blank (uspto)

Result No.	Score	Query		Length	DB	ID	Description
		Match					
C 1	255.4	5.9	7584	3	US-09-446-681-1	Sequence 1, Appli	
C 2	255.4	5.9	7584	3	US-09-446-681-2	Sequence 1, Appli	
C 3	255.4	5.9	7600	3	US-09-469-211A-1	Sequence 1, Appli	
C 4	219.8	5.1	14638	3	US-09-902-540-1106	Sequence 1106, Ap	
C 5	216	5.0	744	3	US-09-902-540-2423	Sequence 2423, Ap	
C 6	184.6	4.2	912	3	US-09-489-039A-6755	Sequence 6755, Ap	
C 7	161.4	3.7	906	3	US-09-489-039A-3950	Sequence 3950, Ap	
C 8	159	3.7	792	3	US-09-252-991A-11222	Sequence 11222, A	
C 9	159	3.7	936	3	US-09-252-991A-11165	Sequence 11165, A	
C 10	156.8	3.6	876	3	US-09-252-991A-12001	Sequence 12001, A	
C 11	155.4	3.6	879	3	US-09-252-991A-11911	Sequence 11911, A	
C 12	150	3.5	798	3	US-09-489-039A-430	Sequence 430, App	
C 13	145.6	3.4	30001	2	US-08-125-468-1	Sequence 1, Appli	
C 14	145.6	3.4	30001	2	US-08-474-933-1	Sequence 1, Appli	
C 15	139.6	3.2	1571	3	US-09-902-540-336	Sequence 336, App	
C 16	139.2	3.2	741	3	US-09-902-540-6232	Sequence 6232, App	
C 17	136	3.1	780	3	US-09-252-991A-12374	Sequence 12374, A	
C 18	136	3.1	954	3	US-09-252-991A-12328	Sequence 12328, A	
C 19	126	2.9	18469	3	US-09-902-540-1205	Sequence 1205, Ap	
C 20	123.6	2.8	579	3	US-09-252-991A-11983	Sequence 11983, A	
C 21	122.4	2.8	5467	3	US-09-902-540-703	Sequence 703, App	
C 22	119.6	2.8	750	3	US-09-902-540-7356	Sequence 7356, Ap	
C 23	119.2	2.7	666	3	US-09-252-991A-12234	Sequence 12234, A	
C 24	118.8	2.7	804	3	US-09-252-991A-15809	Sequence 15809, A	

Db 7344 CGTCTACGCGGAGGCGCGGGTGTAGACACTCGATCGCCCCCTTGGTTCATGGCGTAGTCGA 7285
QY 3317 TGGACTCGGGGAAACCGCGCGCGGGTTCGGGCAGACGAGATGTTGATGATCCGCCGCCGT 3376
Db 7284 TGACGTGCGGGTCTGGCGTAGCGGGCGGATCCGGAGGAAATGTTGACGATGCGACCGCGGT 7225
QY 3377 CGCGCAGTCTGTTTTCAGTCCGTCTGACACAGGAAACAGCGGTGCGCGGACGTTTGACGGCGA 3436
Db 7224 CGTGATCCGGGGCAGGGCATGCCGAGTCACGAGAACCGGGCGCGCTGGTTGATGTCGA 7165
QY 3437 CCAGTCGCTCGAAGACCTCTCTCGTGAATTCCTGATCCGTCGAGCGCTCCGAGCGCTGACGCCG 3496
Db 7164 CCAGACGCTCGAAATCTCTGACAGTGAACGCTCGAGCGCTCCGCGCGGACTGATTCGCG 7105
QY 3497 GCTTGTTCACAGATGTGCAACTCGGGGGGCACTCCGAACTCGCCCATCCCGGGTTCGA 3556
Db 7104 GCTTGTTCACAGGATGTGCAAGCCCTCGGTCTGAGCCCGGAGACCGCTCGAGCGCGGAGTTCGA 7045
QY 3557 ACGCCGCTAGAGCGCGCGCGCTCACCCACAGACGCGAGTTTCGGCCCGGATGGCCAAACG 3616
Db 7044 ACTCCCGCATCAGCTCTTCAGGCCCTCGGGTCGGGACAGGTTCGGCCTTGGACCGGAGCGG 6985
QY 3617 CTTGTCCCGCGTGTCTCGGATGFTCTCAGCGGTCTCTCGCGCGCGCCCTCGCTGCTGC 3676
Db 6984 CGAGSCCCCGGACGCGTGTATGCGCTCGAC-ACCGCAGCGCGCGCGCTCCGATCGGATC 6926
QY 3677 GGTAGTGAATGACGAGCGCCCGTCCGGCGCGCGGCGAGCGGCGATACCGGCTCCGA 3736
Db 6925 GGTAGTGAACGATTAACGGCGGACCGGAGGGGCGCACCGCATCTCGCATGCGCGCGCCCGA 6866
QY 3737 TGCCCCCGCTTCCCGCGTCAACAGGGCG 3765
Db 6865 TGCCCGCGCTGCGCCCGGTGACACGGCG 6837

RESULT 2

US-09-446-681-2
; Sequence 2, Application US/09446681
; Patent No. 6849442
; GENERAL INFORMATION:
; APPLICANT: Archer, John AC
; APPLICANT: Summers, David K
; APPLICANT: Roland, Herve J
; APPLICANT: Powell, Justin AC
; TITLE OF INVENTION: Biosensor materials and methods
; FILE REFERENCE: 0380-P02083-US0
; CURRENT APPLICATION NUMBER: US/09/446,681
; CURRENT FILING DATE: 2000-03-14
; PRIOR APPLICATION NUMBER: PCT/GB98/01893
; PRIOR FILING DATE: 1998-06-29
; PRIOR APPLICATION NUMBER: GB 9713666.7
; PRIOR FILING DATE: 1997-06-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 7584
; TYPE: DNA
; ORGANISM: Rhodococcus corallina
US-09-446-681-2

Query Match 5.9%; Score 255.4; DB 3; Length 7584;

Best Local Similarity 60.6%; Pred. No. 5.3e-35;
Matches 454; Conservative 0; Mismatches 291; Indels 4; Gaps 2;
QY 3020 CGGGTCAGAGATCTTCGCCCTGTGGCATCGACGTACTGGCCGGTGATCCACCGTGAGT 3079
Db 1 CTGGTTAGAGCGGTTGCGCTTGGTGGCGTGCATGACCTGCCCGGTGATCGCACCGCGG 60
QY 3080 COTCGAGGCCGAAAGGCCACACGTCGGCGATGTCGTGGGTCTGCCGA---TGGCGT 3136
Db 61 CGGCGCTGACGAGGAAGGCCACGATCGCGCGATGTCCTCGCGGTGGCGAGTTTTCGA 120
QY 3137 TGAACACGAGTTGGCGGCGAGTGC CGGCGCGCTCGGGGTCTGCGCGCGCGCGTGGT 3196

Db 121 GTGCAGTGGTGTGACGCGCGGTGTGTGGGGCATGCTCGTACCGAAGCCAGTGGCGGT 180
QY 3197 TCATGTCCGTCTCCAGAAACCCGGCGCCACCGGTTGACCGTGTATCCCCGTTCCCCCA 3256
Db 181 TCATGTCCGTATCGAGCGCGCGCCACCGGCTTGGCGGTGATGCTCTGTTGCGCGA 240
QY 3257 GTTCCCTGGCCAGGCGGAGGTGTCACCGCACCTTGGTTCATCGCGTATCCGA 3316
Db 241 CGTCTACGCGAGGCGCGGTGAGCACTCGATCGCCCCCTTGGTTCATGCGGTAGTCGA 300
QY 3317 TGGACTCGGGAAACCGCGCGCGGTTCGGCGACGAGATTTGATGATCCGCCGCCGT 3376
Db 301 TGACGTCCGGTCTCGGTAGCGCGGATCCGGAGGAAATGTTGACGATGCGACCGCGGT 360
QY 3377 CGCGAGTCTGTTTTCAGTCCGTGTCGACAGGAAACAGGGTTCGCCGAGTTCGAGCGGA 3436
Db 361 CGTGCATCCGGGGGAGGGCATGCGGAGTCAACAAAGAACGGGGCGGCTGGTTGATGCGA 420
QY 3437 CCAGTCGCTCGAAGACCTCTCGGTGACTTCCTCGGTGATCCGTCCCGAGCGCGTTCGACGCCG 3496
Db 421 CCAGACGCTCGAAATCTCTGACAGTGAACGCTCGAGGCTCCGGCGCGACTGATTCGG 480
QY 3497 GCTTGTTCACAGGATGTTCGAACTCGGGCGGCACTTCGAACTCGCCCATCCCGCGGTTCGA 3556
Db 481 GCTTGTTCAGCAGGATGTTCGAGCCCTCGGTGAGCCCGGAGACCGTTCGAGCGCGGAGTTCGA 540
QY 3557 ACGCGCGGTAGAGCGCGCGCGCTACCCACGACCGCGAGTTCGGCCCGGATGGCCAAACG 3616
Db 541 ACTCCCGCATCAGTCTTTCAGGCCCTTCGGGTTCGGGACAGGTTCGGCTTCGACCGGCGG 600
QY 3617 CCTGTCCGCGCTGCTCCGATGTCGACGCTCTCTCGCGCGCGCGCTCGCTGCTGC 3676
Db 601 CGAGCCCCCGGACGCGCTGATGCGGTGCAC-ACCGCAGCGCGCGCTTCGATCGGATC 659
QY 3677 GGTAGTGAATGCCACGAGCGCCCCGTCCGGCGCAGCGCGGAGGCGGCCACCGCATCTCGATGCGATGCGCGCCCGA 719
Db 660 GGTAGTGAACGATTACGCGGSCACCGGAGGCGGCCACCGCATCTCGATGCGATGCGCGCCCGA 719
QY 3737 TGCCCCCGCTTCCCGCGGTACACAGGGCG 3765
Db 720 TGCGCGCGCTGGCCCCGCTGACACGCGCG 748

RESULT 3

US-09-469-211A-1/c
; Sequence 1, Application US/09469211A
; Patent No. 6660524
; GENERAL INFORMATION:
; APPLICANT: J. Turck
; APPLICANT: J. Archer
; TITLE OF INVENTION: CONTROL OF GENE EXPRESSION IN EUKARYOTES
; FILE REFERENCE: 9341-021
; CURRENT APPLICATION NUMBER: US/09/469,211A
; CURRENT FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: UK 9828660.2
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 7600
; TYPE: DNA
; ORGANISM: Rhodococcus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (295)..(1035)
; OTHER INFORMATION: ohpR regulator
; NAME/KEY: CDS
; LOCATION: (1261)..(2805)
; OTHER INFORMATION: ohpA transport
; NAME/KEY: CDS
; LOCATION: (2807)..(4720)
; OTHER INFORMATION: ohpB monooxygenase

NAME/KEY: CDS
LOCATION: (5721)..(6665)
OTHER INFORMATION: ohpD catechol 2,3-dioxygenase
US-09-469-211A-1

Query Match
Best Local Similarity 5.9%; Score 255.4; DB 3; Length 7600;
Matches 454; Conservative 0; Mismatches 291; Indels 4; Gaps 2;
3020 CGGGTCAGAGAAATCTTCGGCTGTGGCATCGACACTACTGGCCGGTGATCCACCGTGAGT 3079
7595 CTGGTTAGAGCCGGTTGCGTTGGTGGCTCGATGACCTGCCCGGTGATGCGACCGCGG 7526
3080 CTTGCGAGCCAGAAAGCCACCACTGCGCATGTCTGTCGGTCTGCCGA---TTCGGT 3136
7525 CGCGCTCAGCAGGAAGCCACATCGCGGGGATGCTCTCGCGGTGGCGAGTTTTCGA 7466
3137 TGAACACGAGTTGCGCGCCAGTGC CGCGCGCTCGGGGTCTGCGCGCGCGTGGT 3196
7465 GTGCACTGGTGACCGCGCGGTGGTGGCGGATGTGTCACCGCAAGCCAGTGGCGT 7406
3197 TCATGTCGGTCTCCACGAAACCCCGCGGCCACCGCGTTGACCGTGATCCCGCTTCCCGCA 3256
7405 TCATGTCGGTATCGAGCGCGCGCGGCCACCGCGTTGGCGGTGATGCTGCTGTTCCCGCA 7346
3257 GTTGCCTGCGCCAGGCGGAGCGTGAAGGTTCACCGCACTTGGTCTATCGCGTATCGCA 3316
7345 CGTCTACGCGGAGGCGCGGGTGAACCTCGATCGCGCCCTTGGTCTATGGCGTAGCTGA 7286
3317 TGGACTCGGGGAAACCGCGCGCGGTGCGGCGAGAGATGTTGATGATCGCGCGCGGT 3376
7285 TGACGTGGGTCTGGGTAGCGGGCGGATTCGCGAGAAATGTTGACGATGCGACCGCGT 7226
3377 CGCGCAGTCTGTTTCACTCGTCTGACACGAGAAACAGCGGTGCGCGAGCTTTGACGCGCA 3436
7225 CGTGATCGGGGCGAGGATGCGGATCACGAGAACGCGGCGCTGTTGATGCGCA 7166
3437 CAGTGGTTCGAAGACTTCTCGGTGAATTCGCGTATCGCTCCGAGCGCGTGAAGCGCG 3496
7165 CCAGACGGTCGAAATCTCGACAGTGAACGCGCTCGAGCGCTCCGCGCGGACTGATTCGG 7106
3497 CTTTGTTCACGAGATGCGACTCGGGCGGCACTCCGAACTCGCCCATCCCGGCGTGA 3556
7105 CGTTGTGACGAGATGTCGAGCTTCGCTCGAGCCCGAGACCGTTCGAGCGCGGAGTGA 7046
3557 ACGCCGCTGAGAGCGCGCGGTCAACCAACGCGCGGATTCGCGCGCGGATGCGCAACG 3616
7045 ACTCCGATCAGCTCTTCAGGCGCTTCGCGTGGGACAGGTTCGCGTGGACCGCGCG 6986
3617 CTTGTCCGCGCTGCTCGGATGGTTCGACGGTCTCTCGCGCGCGCGCTTCGCTGCTGC 3676
6985 CGAGGCGCGCGCGCGGATGATGCGTGCAC-ACCGCAGCGCGCGCGTTCGATCGGATC 6927
3677 CTTAGTGAATCCACGAGCGCGCTCGCGCGGCGAGCGGCGGATACCGGTCGA 3736
6926 CTTAGTGAATGATTCGCGCGGACCGGAGCGCGGCGGCGGATCTCGATGCGCGCGCGCA 6867
3737 TCCCGCGGCTTCCCGGCTCACAGGGCG 3765
6866 TCCCGGCTTCCCGGCTGACAGGGCG 6838

RESULT 4
US-09-540-1106/c
Sequence 1106, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B

CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 1106
LENGTH: 14638
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-540-1106

Query Match
Best Local Similarity 5.1%; Score 219.8; DB 3; Length 14638;
Matches 429; Conservative 0; Mismatches 327; Indels 3; Gaps 1;
3009 CCGAAGCGTGGCGGTTCAGAGAAATCGTTCCGCTGTGGCATCGAGTACTGCGCGGTGAT 3068
12645 CTGAGCGCGCGCGCTCAGAGCATCATCCGCCCGAGGCTCGATGCGTGGCCCGTCAAC 12586
3069 CCACCGTGTGTCGTCGAGGCGGCAAGAGCCACCGTCGCGGATGTCGTGGGTCTGCC 3128
12585 CCACCGCTCTCCGCGGACAGCGATGCGACCGCATCCCGGATGTCGTGGGTGACC 12526
3129 GATCGCGTTGAACACGAGGTTGGCGGCGAGTCCCGCGCGCTCGGGGTCTGCCGCG 3188
12525 CACAGCGCGAGCGCAACTGGCGGCCAGCGCTCTGTTTACGTCGCGGTTGTACGAG 12466
3189 CCGTGCCTTC--ATGTCGCTCTCCAGAAACCGCGCGCGCACCGGTTGACCGTATCCC 3245
12465 GACGCGCGCGCGAAGTCCGTCGCGGTCCGCGCGCGCGCGAGCACGTTGACGGAATCCT 12405
3246 CCGTTCGCCCGATGTCCTGCGCGCGGCGGAGTGTGTCGCGACCGTCTGTCAT 3305
12405 CCGCGCGCGCGAGTCTCTTTCGCGAGTACCGGTCGAGCGCTCCAGCGCGCTTTCATGT 12346
3306 CCGTATCCGATGAGTCTCGCGGAAACCGCGCGCGGTTCGCGGCGAGACGATGTTGATGAT 3365
12345 CCGTATGCGCGGTGGCGCGGAGTGAAGCGCGGAGCGCGGAGGATGTTTTCAGGAT 12286
3366 CCGCGCGCGTTCGCGCATGCTGTTTTCAGTCCGTCGTCGAGAACAGCGGTTGCGCGGAC 3425
12285 GCGCGCGCGCATCCGCGAGCACCGCGCTGCGTGAAGAGAAAGTCCGCTTTCAG 12226
3426 GTTGAAGCGGACGATGCTGTCGAGAACCTCTTCGCGTGTGTCGCGTTCGTCGCGGCG 3485
12225 GTGACGCTTCATCAGCGCTGAACTGGCGCTTCGTCGTCGCGAGCTGACTTGGAT 12166
3486 GTGACGCGCGGTGTTTTCACGAGATGTCGAACTCGCGCGGCACTCCGAACTCGCGCAT 3545
12165 GCGATGCGCGGTGTTTTCACGAGAGTTCAGCGCGCTCGCGCGCGGAGTGAAGCGCG 12106
3546 CCGCGCGTTCGAAACGCGCGGTGAGAGCGCGCGCGCTCAACCGAGCGCGGATTCGCGCG 3605
12105 CTCGTCGCGCGCTTCCAGAGCGCGGAGCGCGCGCTGTCGCGCGCGGAGTGAAGCG 12046
3606 GATGCGCAACCGCTGTCGCGCGCTGTCGCGGATGTCGAGGTCCTTCGCGCGCGCGCG 3665
12045 GAGCGCGCGCGCTTTCGCGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 11986
3666 CTCGCTGTCGCTGATGAGTGTGCACTGCCAGCGCGCGCTCCGCGCGCGCGCGCGGAT 3725
11985 CCGCGCGGTGCGGTACGTGAGTATGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 11926
3726 ACCCGGTCCGATGCGCGCGCTTCCCGCGTTCACGAGGCG 3764
11925 GTTCCGCGCGCGTCCAGCGCTTCCCGCGTTCGCGCGCGCGCGCGCGCGCGCGCG 11887

RESULT 5
US-09-540-2423/c
Sequence 2423, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Wiegand, Roger C.
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
 ; FILE REFERENCE: 38-10(115949)B
 ; CURRENT APPLICATION NUMBER: US/09/502,540
 ; CURRENT FILING DATE: 2001-07-10
 ; PRIOR APPLICATION NUMBER: 60/217,883
 ; PRIOR FILING DATE: 2000-07-10
 ; NUMBER OF SEQ ID NOS: 16825
 ; SEQ ID NO 2423
 ; LENGTH: 744
 ; TYPE: DNA
 ; ORGANISM: Myxococcus xanthus
 ; US-09-502-540-2423

Query Match 5.0%; Score 216; DB 3; Length 744;
 Best Local Similarity 56.6%; Pred. No. 2.5e-28;
 Matches 421; Conservative 0; Mismatches 320; Indels 3; Gaps 1;
 QY 3024 TCAGAGATCGTTCCGCTGGGATCGACGTAATGCGCGGTGATCCACCGTGAAGTCGTC 3083
 DB 744 TCAGAGATCATCCGCGCCGAGGCTCGATCGCTGGCCCGTCAACCAAGGCTCTCGG 685
 QY 3084 GGAGCCGAAAGGCCACCACTGCGGATGTCGTCGGGTCTGCGGATGCGGTTGAACAC 3143
 DB 684 CGACAGCAGCATGCGACCGCATCCCGATGTCGTCGGGCTGACCCACAGCGGAGCGC 625
 QY 3144 GGAGTGTGCGGCGAGTGGCGGCGCTCGGGGTCTGCGCGCGGTGCTTC--AT 3200
 DB 624 AACCTGGCGGCGAGCTCTGTTGAGTCTCGGTTGTCAGGAGGAGCGCGCCCGCAA 565
 QY 3201 GTCCGTCTCCAGAAACCCGCGCCACCGCTTGAACGTTGATCCCGGTTCCCGGTTG 3260
 DB 564 GTCCGTGCGGTGCGCGCGCGCGGAGCAGCTTGAAGGAAATCTCCGGGCGCCAGCTC 505
 QY 3261 CTGCGCAGGCGAGCGTGAAGTGTCCACGCACTTGTGATCGGATCGGATCGATCGA 3320
 DB 504 CTTCGCGAGGTACCGGCTGAGCGCTCCACGCGCGCTTCATGTCGCGTAGGCGGCGTG 445
 QY 3321 CTCGGGGAACGCGCGCGGTGCGGCGAGACAGATGTTGATGATCCCGCGCGTCCGCG 3380
 DB 444 GCGAGGAGGTGAAGCGCGGAGCGCGGAGGATGTTGAGATGCGCCCGCCATCCGC 385
 QY 3381 CAGTCGTTTCACTCGCTGCGACAGGAAACAGCGGTGCCCGGAGTTGACGCGGACGAG 3440
 DB 384 CAGCACCGGCAACAGCGCTGCGTGAGGAGAACTGTCCTTGAAGTGAAGTTTCATCAG 325
 QY 3441 TCGGTGGAAGACTCTCTCGGTGATCTTCGATCGTCCGAGCGGCTGACGCGCGGTT 3500
 DB 324 CGCGTCGAACCTGGGCTCGGTCTCGCGGAAGCTGATGAGTGGCCATGCGCGCGTT 265
 QY 3501 GTTACAGGATGTCGAACCTCGGCGGCACTCCGAACCTCGCCATCCGCGGCTCGAAGC 3560
 DB 264 GTTACAGGAGTCCAGCGCTCGCGCGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAG 205
 QY 3561 CGCGTAGAGCGCGCGGCTCACCAACAGCGCGGAGTTGCGCCCGGATGGCCAAAGCGCTG 3620
 DB 204 TTCCAGGAAGCGCGGAGCCCGGTGTCCTCCCACTGAGCGGAGCGGCGCAACCGCTT 145
 QY 3621 TCGCGCGTCTCGGATGTTGTCAGCGGTCTCTCGCGCGCGGCTGCTGCTGCTGCTGTA 3680
 DB 144 GCGCGCTCTGCTTCAATCTGCTTCAACAGCGCGCGCTTCTGTCGCGCGGCTGCTGTA 85
 QY 3681 GTGAGTCCGAGCGCGCGCTCGCGCGGAGCGGAGGAGTACCGCGTCCGATGCC 3740
 DB 84 GTCAGGATGATGCGGTGCGCGCGCGGCGGAGCTTGAAGCGCATGTTCCGCGCGAGTCC 25
 QY 3741 CCGGCTTCCCGGCTCACAGGCG 3764
 DB 24 ACGGCTTCCCGGCTGATGAGAGC 1

RESULT 6
 US-09-489-039A-6755/C
 ; Sequence 6755, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 6755
 ; LENGTH: 912
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 ; US-09-489-039A-6755

Query Match 4.2%; Score 184.6; DB 3; Length 912;
 Best Local Similarity 53.5%; Pred. No. 6.2e-23;
 Matches 410; Conservative 0; Mismatches 354; Indels 3; Gaps 1;
 QY 3024 TCAGAGATCGTTCCGCTGGGATCGACGTAATGCGCGGTGATCCACCGTGAAGTCGTC 3083
 DB 912 TCAGATGTTGTTCCGCGCGGACCTCGATCCGCTGGCGGTACCCAGCGGTTATCGTC 853
 QY 3084 GGAGCCGAGAAAGGCCACCACTGCGGATGTCGTCGGGTCTGCGGATGCGGTTGAACAC 3143
 DB 852 GCGTAGCAGGCTGCGGATCATCGGCCCAATATCTCGGGAACGCCACCCCTCCAGCGC 793
 QY 3144 GGAATTGCGGCGAGTGGCGGCGCTCGGGGTCTGCGG---CGCGGTGCTTCAT 3200
 DB 792 CGTCATGCGCGCAAACTCGCGCTTACCTCTGCGTCTCGCGCACCAAGCCACCGCCAAA 733
 QY 3201 GTCCGTCTCCACGAAACCGGCGCCACCGGTTACCGTGTACCGTATCCCGTTCGCCAGTTG 3260
 DB 732 ATCGTGGCGATAGCGCGGCGGCGATGATTAAGCGTGTGATGATCGCGCGCGCGAGTC 673
 QY 3261 CCTGCGCAGGCGAGCGTGAAGCTGTCCACCGCACCTTGTGTATCGGATATCGATGGA 3320
 DB 672 GCGGCCATGTACAGCTCAGCATCTCCACAGCGCTTTGGCGGCGGTAGGCGGAGAA 613
 QY 3321 CTCGGGAAACGCGCGCGGTGCGGCGAGCAGATGTTGATGATCCGCGCGCGTCCGCG 3380
 DB 612 GCCAGGATAGGATACGCGGCTCAGCCCGGAGGAAAGTTAACGATCCGCGCGCGTCTGC 553
 QY 3381 CAGTCGTTTCACTCGTCTGAGCAGGAAACAGCGGTCCCGGACGTTGAGCGGCGACCGAG 3440
 DB 552 CAGCAGCGGAGAGGTTTGTACAGGAAATAACCCCTTAACTGAGAGTTAAACAG 493
 QY 3441 TCGGTGGAAGACCTCTCGGTGACTTCGTCGATCCGTCGCGCGGCTGACGCGCGGCTT 3500
 DB 492 CCCGTGCAACTCGCTTCGCGGTTTCGCGCAGCGGCGGAAATTCGCGTCCGCTGCGTT 433
 QY 3501 GTTCAACAGATGTCGAACCTCGGCGGCACTCCGAACTCGCCCATCCCGGCGTCCAGCGC 3560
 DB 432 ATTGATCAGGTGTCGAAAGTGGCCCGCCCGCCAGACGCTGGCCAGCGGCTGCGAGGCT 373
 QY 3561 GCGCTAGAGCGCGCGCTCACCAACAGCGCGGATTCGCGCGGATGGCCAAACGCGCTG 3620
 DB 372 GTCGCGGAGGCGCGGAGCTGCGGCTTGTGCGCATATTCAGCGGCGAGGCGATGGCCTT 313
 QY 3621 TCCGCGCTGCTCGGATGTTCTCGACGTTCTTGGCGCGCGCTCGCTGCTGCTGCTGTA 3680
 DB 312 GCGGCCAGGCGCGGATGTCGCGGACCACTGCTCGGCTTTCGAGACTGCGCGGTA 253
 QY 3681 GTGAGTCCCAAGAGCGCGCGTCCGCGGCGAGCGGAGGCGGATACCGGCTCCGATGCC 3740
 DB 252 GGTGACGATGACATCCCGCCCGCGGCGATATCCAGTCTGCGGTCTGACGCCAAGGCC 193

Db 426 TCCTGATCACCAGCGCGCGTCCGCTTGTGATCGCGCGCTTGGTGGGACGAAGA 485

Qy 3416 GTGCCCGGACGTTGACGGCGACAGTCGGTCAAGACCTCTCTCGGTGACTTCGGTGATCC 3475

Db 486 CGCTGGCAGCTTGTATCGCAGGCTCGCTCGAAGTCGGCCAGCGGACGCTCTTCCACCG 545

Qy 3476 GTCCGAGCGCTGACCGCGCGTGTGTTTACACAGGATGTGAACTCGGGCGGCACTCCGA 3535

Db 546 AACCGAGGCGAGGACCGCGCGTGTGTTGACAGGATGTGATGCG-----TCCGA 596

Qy 3536 ACTCGCCCATCCGCGGTGCAAGCGCGTGTGAGCGCGCGCTCACCCAGACGCGGA 3595

Db 597 AGCGAGCGCCCTTGTGTCAGCGCGCGACGAGCGCGCGCGCT-----CGGCG 647

Qy 3596 GTTCGCGCGATGCGCAACGCGCTGTCCCGCGCTGTCTCGGATGTCTCGACGCTCTCTC 3655

Db 648 TATCGCATGACGCGCAGGATGCTGGCTCCGCGCGCTTCGAGTTCCGCGCGCAGGCGCC 707

Qy 3656 GCGCGCGCGCTGCTGTGCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 3715

Db 708 GCGACTTGTCCGCGAGCTGAGTGTGAGGCGGACGCTGTGCGCGCTTCGCGCGCGCGCG 767

Qy 3716 GCAGGCGGATACGCGTCCGATGCGCGCGCTTCCCGCGCTCACAGGCGGCTTTCGCGCT 3775

Db 768 GCGCAACGATGCG 827

Qy 3776 CCAGCGGT 3783

Db 828 CAGGGAT 835

RESULT 11

US-09-252-991A-11911/c

; Sequence 11911, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 11911

; LENGTH: 879

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-11911

Query Match 3.6%; Score 155.4; DB 3; Length 879;

Best Local Similarity 54.6%; Pred. No. 6.2e-18;

Matches 387; Conservative 0; Mismatches 301; Indels 21; Gaps 3;

Qy 3078 GTCTCGGAGGCGCAGAAAGCCACACGTCGGCGATGTCTCGGGTCTGCGGATCGGTT 3137

Db 834 GTAGGCGCTTCCGGCGCGCGCAAGTAGGCGAAGCTGGCAATTCCTCGCTGCGGCG 775

Qy 3138 GAACAGGAGTTTGGCGCGCAGTGTCCGCGCGCTTGGGGGTCTGCGCGCGCGCTCGTT 3197

Db 774 GTAGCGCGCAGCGCCATCAGGCGCTTGGGGCTTGGCGAAGTCTGCGCTGCTCGGTT 715

Qy 3198 CATGTCCGTCTCACGAAACCCGCGCGCACCGGTGTGACCGTGTATCCCGGTTCCCGCAG 3257

Db 714 CATGTCCGTATCCACGAGACCCGCGCTGACGTTGTTTACCGTGTATCCCGCGCGCGCGAG 655

Qy 3258 TTGCTTGGCGCGCGCGCGGTGTGACGCGTGTCCACGCGACCGCTTGGTGTATCGCGTATCCGAT 3317

Db 654 GTCCGCGCGCGCGCGCTTGGTTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 595

Qy 3318 GGACTCGGGGACG---CGCGCGCGGTCCGCGACAGAGATGTTGATGATCCGCGCGCC 3374

Db 594 GCGCGCGGAAACGCGATGCGTTTCGGCAATTGCTGCGGATGCTGATCACCAGCGCGCC 535

Qy 3375 GTCCGCGAGTGTGTTTTCAGTTCGTGTGACACGAGGAAACGCGGTGCGGACGTTGACGCG 3434

Db 534 GTCCGCGCATGTGTTGACCGCGCGCTTGGTGGCGACGAGACGCTTGCACGTTGATCGC 475

Qy 3435 GACCAAGTCGTTGCAAGACCTCTCTCGGTGACTTTCGTGTATCCGTCGAGCGCTGACGCG 3494

Db 474 CAGGCTCCGCTCGAAGTCGCGCAGCGGAGCTTCTTCCACCGAACCAGAGGCGGAGGACGCG 415

Qy 3495 CCGCTTGTTCACGAGATGTGAACTCGGGCGGCACTCCGAACTCGCCCATCCCGCGCTC 3554

Db 414 GCGCTTGTGTGACGAGATGTGATGCG-----TCCGAAGCGAGCGCGCTTCGTC 364

Qy 3555 GAAACCGCGGTAGAGCGCGCGCGCTCACCCACGACGCGGAGTTTCGCGCGGATGCGCAA 3614

Db 363 GACGCGGCGGACGAGCGCGCGCGCT-----CGGCGCTATCGGCATGACGAGGCGCAG 313

Qy 3615 CCGCTGTCCGCGCTGCTTCGGATGCTTCGACGCTTCTTCGCGCGCGCGCTTCGCTGCT 3674

Db 312 GATGCTGCTCCGCGCGCTTTCGAGTTTCGCGCGCGCGGCGCTTTCGCGCGAGCT 253

Qy 3675 GCGCTAGTGGACTGCCAGGACGCGCGCTTCGCGCGCGCGCGGCGATACCGCGTCC 3734

Db 252 GACGTAGTGAAGCGACGCTGCGCGCTTTCGCGCGCGCGCGGCGAAGCGATGCGCGCGCC 193

Qy 3735 GATGCGCGCGCTTCCCGCGCTCACGAGGCGCGCTTTCGCGCGCGCGCTTCGCGCGT 3783

Db 192 GATGCGCGGGAACCGCGCTTTCGCGCGCGCGCTTTCGCGCGCGGAT 144

RESULT 12

US-09-489-039A-430/c

; Sequence 430, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 430

; LENGTH: 798

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-430

Query Match 3.5%; Score 150; DB 3; Length 798;

Best Local Similarity 51.9%; Pred. No. 5.2e-17;

Matches 389; Conservative 0; Mismatches 355; Indels 6; Gaps 2;

Qy 3024 TCAGAGAAATCGTTTCGCGCTGTGCGATCGACGTACTGCGCGCGGTGATCCAGTGAAGTCGTC 3083

Db 798 TCAGATGCGCAATTCGCGCGGACACTTCAATGCGTGGGCAATTCACCGAGGATGATCATC 739

Qy 3084 GGAGGCGCAGAAAGCCACACGTCGCGGATGTGTCGGTCTGCGGATGCGGTTGGAACAC 3143

Db 738 GGATTAACAGCGGCGCATCATCTGCGCGGATATCTCTGCGAGCGCGCGCGCGCGCG 679

Qy 3144 GGAGTTGGCGCGCAGTCCGCGCGCGCTTCGCGGCTGCGG---CGCGCGTGGTTCAT 3200

Db 678 CGTATCGCAGCAGCAGCAGTTCGCGGCTTATCCCGCACACACCGCGCGCTGAA 619

Qy 3201 GTCCGTCTCCACAAAACCG 3260

Db 618 ATCGTGGCGATGGCG 559


```

3261 CCTGCCAGGCGAGCGTGTCTCCACCGCACCTTGTGTCATCGGTATCCGATGGA 3320
Db      |||||
558  AAACGCCATATAGCGGTTCAGGTTTCCACCGCGGATTCATGATCGATAGGGGCGCG 499
Qy      |||||
3321 CTCGGGGAAACCGCGCGCGGTTCGGGCGACAGAGATGTTGATGATCGCGCGCGGTGCG 3380
Db      |||||
498  GTTCGCCATCAGATGCGGGTGAGSCGGAGCAAAATTCATCTATTGCGCGCGCATCGCG 439
Qy      |||||
3381 CAGTCGTTTTCAGTCTGCTGGACAGGAAACAGCGGTGCCCGGAGCTTGAACGGGACAG 3440
Db      |||||
438  CAGCAGCGGCAATAGTTTTCGCTGAGGAGAAACAGCTCTTCAAGTGGATGCGATACAG 379
Qy      |||||
3441 TCGGTGGAAGACCTTCTCGGTGACTTCCGTGATCGTCCCGAGCGCTGACGCCCGGCTT 3500
Db      |||||
378  CGCATCGAATTCGGTTCGTTGGCGTTAAGAAATCCCATGCGCTGGCGTTGCGCGCATT 319
Qy      |||||
3501 GTTCACCAAGATGTCGAATCTCGGGCGGCACTCCGAACCTCGCCCATCCGGCGTTCGAACGC 3560
Db      |||||
318  ATTGACCAGATAATCAAAATTTATCG---CGCCAGCTCCGTGAGCGCGCTGAAAAGC 262
Qy      |||||
3561 CGCGTAGAGCGCGCGCGTCAACACAGACGCGAGTTTCGGCCCGGATGGCAACGCTG 3620
Db      |||||
261  GCTAACAAATGCTCAAAAGTGTGCAATTCGCGCGCGCTGAAAGGCAACGCGATCGCGCG 202
Qy      |||||
3621 TCCCGCGTCTCGGATGCTCGACGCTCTCGCGCGCGCGCTCGCTGCTGCGCGTA 3680
Db      |||||
201  GCGCGCCAGTCTCAACCCGTTGTAACCACTGCTTCGCTTCGCGCAAGACGCGTTTGTGA 142
Qy      |||||
3681 GTGACTGCCACGAGCGCGCGTTCGCGCGCGCAGCGCAGGCGGATACCGGTTCGATGCC 3740
Db      |||||
141  GGTCACTACTACATTTACGCGCGCGTGTGCGAGCGCTTCGACGCGTTCGCGCGTCCAGGCC 82
Qy      |||||
3741 CCGGTTCCCGCGTCAACAGGCGGTCTT 3770
Db      |||||
81  CCGGTCGCGCGTGTCAACAGGCGGATTTT 52

```

```

RESULT 13
US-08-125-468-1
; Sequence 1, Application US/08125468
; Patent No. 5589385
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy
; APPLICANT: Fantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlortetracycline and tetracycline Formation and cosmids
; TITLE OF INVENTION: useful therein
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESS: American Cyanamid Company
; STREET: One Cyanamid Plaza
; CITY: Wayne
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07470
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/125,468
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsevdos, Estelle J
; REGISTRATION NUMBER: 31,145
; REFERENCE/DOCKET NUMBER: 31,255-02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (201)831-3241
; TELEFAX: (201)831-3305

```

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 30001 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

Query Match      3.4%; Score 145.6; DB 2; Length 30001;
Best Local Similarity 53.9%; Pred. No. 3.8e-16;
Matches 348; Conservative 0; Mismatches 289; Indels 9; Gaps 2;

Qy      3128 CGATGCGGTTCGAACACGAGTTGGCGGCCAGTGCAGCGCGCGCTCTCGGGGTCTGCGGCC 3187
Db      |||||
7734  CGAAGCGGAGACGATCTCTCTCGGTGTCGCCAGGTGCGCGCGTACGCTTGGCGGA 7793
Qy      3188 GCGGTGCGTTTCATGTCGTCTCCACGAAACCGCGCGCGCACCGGTGACCGGTGATGCC 3247
Db      |||||
7794  CCGCACGGCCATCGGCGTCTCGACGTAGCCGGGCGACAGCGCGTTGACCGTGTGCGCG 7853
Qy      3248 GTTCCCGCAGTTGCTGCGCAGGCGGAGCGTGTCCACCGCACCTTGTGTCATCG 3307
Db      |||||
7854  TGTGGCGAGTTCTTGTGCGCAGCGCTTGTGTAAGCGATGACCGCGCTTGGAGGCG 7913
Qy      3308 CGTATCGATGACTTCGGGGAACCGCGCGCGGTTCGGGCGACAGATGTTGATGATCC 3367
Db      |||||
7914  AGTAGGGGCGCCACGCGGACACCTGCTGTCGCCCGGTGAGCGCGATGATGATCC 7973
Qy      3368 GCCCGCGTTCGCGCAGTGTGTTTCAGTCCGTGCTGCAACAGGAAACAGCGGTGCCG 3427
Db      |||||
7974  GCCCGTCCCGCGCGCTCCATGTCGCGCGGTGTGAGGACCTCGCGGTGACGCGGGAAGA 8033
Qy      3428 TGACGGGACAGTCGCTCGGAGACCTCTCGGTGATCTTCGTCGATCGCTCCGAGCGCG 3487
Db      |||||
8034  CGCTGCTGAGTTGCTGTCGATCAGCTCTCCACAGCTCTCGGTGAGGTGAGGTGA 8093
Qy      3488 TGACCGCGCGTGTTCACAGGATGTCGAACCTCGGGCGGCACTCCGAACCTCGCCCATCC 3547
Db      |||||
8094  CGCCACCGCGTTCGGTTCGGCGTGTGTA-----CCAGCAGCGCGATCGCGCGGAAGC 8147
Qy      3548 CGCGTTCGAACCGCGCTGAGAGCGCGCGCTCACCCACAGCGCGAGTTGCGGCCCGGA 3607
Db      |||||
8148  GGTCCACCGCGCGGACGAGCTGCTACGTCCGCGCGGAGCGGACGTCGCGCGCGGA 8207
Qy      3608 TGGCCACGCGCTGTCGCGCGTCTCGGATGCTCTGACGCTCTCTCGCGCGCGCGCT 3667
Db      |||||
8208  GGCGTCCACCTCCAGGCGCTTCGCGCGGAGCGGCGGACCGTCTCGCGGACGCGCTCGG 8267
Qy      3668 CGCTGCTGCGTGTGACTGCCACAGCGCGCTCGCGCGCGGCGCGGCGCGGCGGATAC 3727
Db      |||||
8268  CGGTGCGGCGCAGATGAG---ACGCTACGCGCGCGCGGCGGCGCGCTCGGCGCTGG 8324
Qy      3728 CGCGTCCGATCCCGCGCTTCCCGGTTCCACAGGCGGCTTTGCC 3773
Db      |||||
8325  CCAGTCCGATCGCGTGTGTCGCGCGGTCACACAGGCGGACGTCGCG 8370

```

```

RESULT 14
US-08-474-933-1
; Sequence 1, Application US/08474933
; Patent No. 5866410
; GENERAL INFORMATION:
; APPLICANT: Ryan, Michael J.
; APPLICANT: Lotvin, Jason A.
; APPLICANT: Strathy, Nancy
; APPLICANT: Fantini, Susan E.
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for
; TITLE OF INVENTION: chlortetracycline and tetracycline Formation and cosmids
; TITLE OF INVENTION: useful therein
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESS: American Cyanamid Company

```


STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: USA
ZIP: 07470
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,933
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125,468
FILING DATE: 22-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Tevdos, Estelle J
REGISTRATION NUMBER: 31,145
REFERENCE/DOCKET NUMBER: 31,255-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3241
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30001 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match 3.4%; Score 145.6; DB 2; Length 30001;
Best Local Similarity 53.9%; Pred. No. 3.8e-16;
Matches 348; Conservative 0; Mismatches 289; Indels 9; Gaps 2;
3128 CGATGCGGTGTAACACGAGTGTGGCGCCAGTGTCCCGCCGAGTGTCCCGCCGCTCGGGGTCTGCGGCC 3187
Db CGAGCGCGAGGACGCTCTCTCGTGTGTGCGCCAGGTGCGCGGTACGCGTGTGCGGA 7793
3188 GCGGTGCGTTCATGCTGCTCAAGAAACCGCGCCGACCGGTTGACCGTGTGACCGTCC 3247
Db CCGCACGCGCATCGGCGTCTGACGTAGCGGGGAGACGCGGTGTGACCGGTGTGCGG 7853
3248 GTTCCCGCAGTGTGCGCGAGCGGTGAGCGTGTCCACCGCACCTTTGTCATCG 3307
Db TGTGGCGAGTTCCTTGGCGAGCGCTTGTGAGCGGTGAGCGCGCTTGGAGCGG 7913
3308 CGTATCCGATGGAATCGCGGAAACGCGCGCGGTGCGCGGACGAGATGTTGATGATCC 3367
Db AGTAGGGGCGCCAGCGGACACCTGCTTGGCGCGGTGGAGGCGACGCTGATGATCC 7973
3368 GCGCGCGTGGCGAGTGTTCAGTTCGCTGTGACACGAGAACAGCGGTGCGCGACGT 3427
Db GCGCGTGGCGCGCGCTCCATGCCCGGTGTGAGGACCTCGCGGTGACGCGGAAGA 8033
3428 TGAAGCGACCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3487
Db CGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 8093
3488 TGAAGCGCGGTGTGTTTCAAGAGATGTGAACTCGCGCGCGCACTCGCAACTCGCCCATCC 3547
Db CGCCACCGCGTGTGCGTGGCGTGTGTTGA-----CCAGCAGCGCGTGTGAGTGTG 8147
3548 CGCGGTGCAACCGCGGTGAGCGCGCGCGTGTGACCGAGCGCGGTGTGCGCGCGGA 3607
Db GGTTCACCGCGCGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 8207
3608 TGGCCAAACGCTGTGCGCGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3667
Db GGCCTGTCCACTCCAGCGCGCTGCGCGAGCGCGGTGTGAGTGTGAGTGTGAGTGTG 8267

3668 CGCTGCTGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 3727
Db CGGTGCGCGCGAGATGAAG---ACGCTCAGCCCCCGCGCGCGCGCTCGCGCTGG 8324
3728 CGCGTCCGATGCGCGCGCTTCCCGCGTCCACGAGCGGTCTTTGCC 3773
Db CCAGTCCGATGCGCGTGTGCGCGCGTCCACGAGCGGTGCGCG 8370

RESULT 15

US-09-902-540-336/c
; Sequence 336, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 336
; LENGTH: 1571
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-336

Query Match 3.2%; Score 139.6; DB 3; Length 1571;
Best Local Similarity 53.0%; Pred. No. 3.3e-15;
Matches 410; Conservative 0; Mismatches 334; Indels 30; Gaps 4;
3037 CCGCTGTGGCATCGACGTACTGCGCGGTGATCCACCGTGTGCTGCGGAGCGCAAGAG 3096
Db CACCATCGACCGCATGAGTGTCTGTGATGATGCGCCGCTGCGCATCGCCAGGTGC 1368
3097 GCCACACGCTCGCGCATGCTGCGGTGCGCATGCGGTGGAACACGAGTTGGCGGCC 3156
Db CCGACGCTGGCGCATGCTCTCGGCTGCGGTAGTGTAA-----GGGACATG 1317
3157 AGTGGCGGCGCGCTCGGGGTCTGCGCGCGCTGCGGTTCATGTCCTCTCCAGAA 3216
Db AGCGCCAACTCGCGCGCGACGCGCGCGCTGCGCGGTTCATGTCGTATCCGTCGAC 1257
3217 CCGGCGCGCACCGCTGACCGGTGATCCCGCTTCCCGCAGTTGCTGCGCAGGCGGAGC 3276
Db CCGGGTGTGATCAGCTTCTGTGATGCGCGCTGCGCATCTCCGTCGCGAGCCCTTC 1197
3277 GTGAGCGTGTCCACCGCACCTTGTGTCATCGGTATCCGATGAGTCTG---GGGAACGG 3333
Db GTGAGCGCATGACGCGGACTTGTCTCATGATGAGCGTGTGACTCGCGCATAGGCGACG 1137
3334 CCGCGGCTCGCGGACGAGATGTTGATGATCGCGCGCGCTGCGCGAGTGTGTTTCACT 3393
Db CGTGGCGGAAAGCAGAAACCAATCGAGATGATGCGCGCACTCGCGCATGTGCGGAGC 1077
3394 CCGTGTGACGACGAGAACGCGTGTGCGCGCGCGCATGCGCGCGCGCATGCGTGTGAGAGCC 3453
Db GCGGTGTGAGACGCGACGAAACGCGCTCGAACTGTGATTTGCCAGCGCTCTGTGAGTTC 1017
3454 TCTCGGTGACTTTCGTGTATCGTCCGAGCGCTGACGCCCGCGTGTGTTTACAGAGATG 3513
Db TCGAGTGTGACCTTTTCGAAACGCTCGGAAGGAGATGCGCGCGTGTGTTGACAGAGATG 957
3514 TCGAACTCGCGCGGACACTCGGCACTCGCCCATCCCGCGCTGCAACCGCGGTGAGCGCG 3573
Db TCCAGCGCG-----CCCAACTCGGACGCGCGCGTGTGACTGTGCGCGCATGCGCGG 906
3574 GCGCGTCAACCGACGCGCGAGTTTGGCGCGCGAGTGTGCGCGCGCGCGCGCTGTGCGCGCTGCTC 3633

Db	905	TCGGGGTCTCCGTTG-----TCCGCCCTGGATGGCCACGGCCCCGTCGGCCGAGTGCC	855
Qy	3634	CGGATGGTCTCGACGGTCTCTCGCGCCGCCGCTCGCTGCTGCGCTAGTGGACTGCCACG	3693
Db	854	TTGAGTTCAACCGACCCAGCGTGGCGCTTTGTGGCCAGAGTTCCGCATAGGTGAAGGCTACG	795
Qy	3694	AGCGCCCGTCCGGCGCCAGCCGCGAGGGCGATACCGGTCGATGCCCGCGCTTCCCCCG	3753
Db	794	TCCGCCCTCCCTTTGCCAGATGCCGAACGATGGCGCGCGGATGCCACACGCGTGCCACCT	735
Qy	3754	GTCACCGGGCGGCTTTGCCCTCCAGCGGTCTTCCATACCTCGTCCCATGTGCA	3807
Db	734	GTCACCAACGGCGCTTGCCACTGAACGCCATCATGTCCACCTCTACTTTGGA	681

Search completed: January 19, 2006, 03:06:23
Job time : 711 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2006, 09:46:57 ; Search time 15841 Seconds
(without alignments)
12836.100 Million cell updates/sec

Title: US-10-017-471B-19
Perfect score: 4346
Sequence: 1 gtcgacgacggtcggtt.....gcgtagagcgcgggtacc 4346

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gsl1:*
10: gb_gsl2:*
11: gb_gsl3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	150.6	3.5	837	9	CC620734	OGQAC91TM
2	150.6	3.5	870	9	CC629147	OGUAC20TC
3	148.4	3.4	917	10	CG096128	PUFZK93TD
4	146.8	3.4	864	10	CL963775	OGIFCC038
5	145.4	3.3	582	6	CB854086	UI-CF-DU1
6	136.8	3.1	1083	7	CK214283	FGAS02620
7	136.2	3.1	1257	10	CL948439	OGIFSB005
8	135.6	3.1	843	7	CN823495	Oa_splbn
9	135.4	3.1	800	7	CN145046	WOUNDI_26
10	134.2	3.1	825	10	CZ544259	SRAA-aad5
11	134.2	3.1	890	8	DR887252	JGI_CUNC6
12	133.8	3.1	764	7	CN130030	RHOH1_38
13	132	3.0	780	7	CO202814	Oa_splbn
14	132	3.0	809	9	CC398497	FUHLIC28TD
15	131.6	3.0	795	9	BZ537000	OGAGU62TC
16	131.2	3.0	845	8	DN258874	Meso03087
17	130.6	3.0	2332	10	AG363333	Mus muscu
18	130	3.0	767	10	CG333497	OGIEE86TH
19	128.6	3.0	674	6	CD866271	AZO2.102P
20	127.4	2.9	750	10	CL982611	OGIFSC048
21	127.2	2.9	730	6	CA140180	SCEZRT202
22	126.2	2.9	769	7	CO423399	GGEZHT100

23	126.2	2.9	2332	10	AG363333	Mus muscu
24	125.8	2.9	798	10	CL974024	OGIFCC042
25	125	2.9	696	6	CD866272	AZO2.102P
26	125	2.9	1081	7	CK211362	FGAS02320
27	123.8	2.8	664	8	DN483106	root3_11
28	123.2	2.8	1885	10	AG363563	Mus muscu
29	122.6	2.8	697	6	CA130420	SCGRT100
30	121.8	2.8	782	6	CA258734	SCGRT301
31	121.4	2.8	767	9	CC670424	OGWIY21TH
32	121	2.8	839	8	DR027220	bda010012
33	120.6	2.8	1569	10	AG341503	Mus muscu
34	118.8	2.7	841	10	CG096126	PUFZK93TB
35	118	2.7	709	1	AJ702939	AJ702939
36	117.2	2.7	795	10	CL958605	OGIFCC001
37	117.2	2.7	828	9	CC670432	OGWIY21TH
38	117	2.7	861	6	CA104193	SCJFHR103
39	117	2.7	1569	10	AG341503	Mus muscu
40	116.6	2.7	778	6	CF637158	zmrw00_0
41	115.8	2.7	550	7	CN822865	Oa_splbn
42	115.8	2.7	823	7	CN822937	Oa_splbn
43	115	2.6	1821	10	CL090560	ISB1-17N1
44	114.8	2.6	461	8	CX949662	UMC-bcl_0
45	114.6	2.6	1821	10	CL090560	ISB1-17N1

ALIGNMENTS

RESULT 1
LOCUS CC620734 837 bp DNA linear GSS 19-JUN-2003
DEFINITION OGQAC91TM ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0364P13,
genomic survey sequence.
ACCESSION CC620734.1 GI:31987155
VERSION GSS.
KEYWORDS Zea mays
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 837)
AUTHORS White, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Renick, A., Frazer, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nurnberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
TITLE Consortium for Maize Genomics
JOURNAL
COMMENT Other GSSs: OGQAC91TC
Contact: Cathy Whiteaw
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: methylation filtered.
Location/Qualifiers
1. 837
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0364P13"
/clone_lib="ZM 0.7 1.5 KB"
/note="Vector: pBCSK; Site 1: HincII: 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN

Query Match 3.5%; Score 150.6; DB 9; Length 837;
Best Local Similarity 52.3%; Pred. No. 4.6e-18;
Matches 417; Conservative 0; Mismatches 359; Indels 21; Gaps 3;
QY 3037 CGCCTGTGGCATCGACGTACTGGCGGTGATCCACCGTGTGTCGGAGCCAGAAAG 3096

Db 160 ACGCCGCGCCACGGGTGGGAGTTGGGAGGACACCTTCGGCGCCGATGCCCGCGGAGCCG 101

Qy 3751 CCGGTACCAAGGGCGGTCTTCCCTCCAGCGGTCTTCATACCTCGTCCCATGTGCACGC 3810

Db 100 CCACTGATACGCGTACGCGCGCGTTCGAGCGGGAGCAGCAGCAGCGTTCCTCGGCGCTT 41

Qy 3811 ATATCAGCCCCCGCGT 3827

Db 40 GCGACTGCCATGGCTGT 24

RESULT 3

CG096128 917 bp DNA linear GSS 20-AUG-2003

LOCUS PUFZK93TD ZM 0.6.1.0 KB Zea mays genomic clone ZMMBta0768017,

DEFINITION genomic survey sequence.

ACCESSION CG096128

VERSION CG096128.1 GI:33978422

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 917)

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.

TITLE Maize Genomics Consortium

JOURNAL Unpublished (2003)

COMMENT Other GSSs: PUFZK93TB

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: Tg

Class: sheared ends.

FEATURES

source : 1..917

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBta0768017"

/clone_lib="ZM 0.6.1.0 KB"

/note="vector: PCR4-ROPO; Site 1: EcoRI; 0.6-1.0 kb high Cot selected genomic DNA library"

ORIGIN

Query Match 3.4%; Score 148.4; DB 10; Length 917;

Best Local Similarity 54.3%; Pred. No. 1.2e-17; Indels 33; Gaps 4;

Matches 408; Conservative 0; Mismatches 311;

Qy 3037 CCGCCTGTGGCATCGACGTACTGGCCGCTGATCCACCGTGAAGTTCGAGTCCGAGGCCAGAAAG 3096

Db 192 CCGCGGTGGCGCGATGACCTGCGCGTTGACCCACTCGGCGGCTCGGTGCACAGGAAG 251

Qy 3097 GCCACCAAGTCGGCGATGTCGTTCGGGTCTGCCGATGCGGTTGAACACGAGATTGGCGGCC 3156

Db 252 CCGACCACCGCGCGCATGTCACCGGCTCGCCGA---GCCTCTGCATCGGGTTTGTCTCC 308

Qy 3157 AGTGC CGGCGCGCTCGGGGTCTGCCGCGCGGTTCGATGTCGTTCGATCCACGAAA 3216

Db 309 ACGGCGTGTCCACCAATGGCCCTCGCTTTGCCCTTGAAGAACATGTCGTGCGCGTGGCC 368

Qy 3217 CCCGCGCACCGGTTTACCGTGTATCCCGGTTCCCGGTTCCCGGTTGCTGGCCAGGCGAGC 3276

Db 369 CCGGCGCGACGAGTTGGCGTGTACGCGGGTGGCCCGAGCTCCTTGGCCATGGTCCGC 428

Qy 3277 GTGACGCTGTCCACCGCACCCCTTGGTTCATTCGCGTATCCGATGGAATCGGGAAACGCGCGC 3336

3404 CCAGGAACGCGGTGCCCGGACGTTGACGGCGACAGTCCGTGCGAAGACCTCCTCGGTGA 3463
|||||
Db 222 TGAGGAAACCGCGCCTTTGAAGTGAATGGCCACCACTTTGTGCAACTGCTCTTCGGTGG 163
3464 CTTTCGGTATCGTCCCGAGCGCGTACGCCCGCGTGTGTTTACCAAGGATGTGCAACTCGG 3523
Db 162 TTTTCGGCAAGTTGCGTGAGCGCGCGATGCGCGGTGTTGTATCAGGAATTAATGTT 103
3524 GCGGCACTCCGAACCTGCCATCCCGGCGTGAACCGCGCGTAGAGCGCGCGCGGTAC 3583
Db 102 GCTGGCGGAAACGCTTTTGAGCACCGTAGCGATTGTCGCCGACAAAGCATCGAAGTTGC 43
3584 CCACGACGCGCGATTC 3599
Db 42 CGCTCTGGCTGACGTC 27

RESULT 6
CK214283/c
LOCUS
DEFINITION
FGAS026206 Triticum aestivum FGAS: Library 6 CAP GATE 1 Triticum
aestivum cDNA, mRNA sequence.
ACCESSION
CK214283
VERSION
CK214283.1 GI:39620387
KEYWORDS
EST.
ORGANISM
Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 1083)
Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,
Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,
Penniket, C., Roach, J.L. and Sarhan, F.
Functional Genomics of Abiotic Stress in Wheat and Canola Crops
Unpublished (2003)
Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas.ests@usask.ca
This sequence is the direct result of the Base calling software
Phred (default parameters). It is the raw base calls. To aid in the
identification of the high quality insert the software Lucy
(default parameters) has been run on this sequence. Lucy identified
the region [66,721].
Plate: L6B012 row: F column: 03.

FEATURES
source
1. .1083
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone_lib="Triticum aestivum FGAS: Library 6 CAP GATE 1"
/note="Organ: Crown and leaf; Vector: pCMV.SPORT6; Crown
(50%) and leaf (50%) tissues from wheat cultivar Norstar
after short exposure times to low temperature in the light
and in the dark. 12 mRNA populations were combined before
constructing the library. The first 6 populations: After 7
days of growth at 20C from wheat cultivar Norstar after
short exposure times to low temperature in the light and
in the dark. 12 mRNA populations were combined before
constructing the library. The first 6 populations: After 7
days of growth at 20, wheat plants were transferred to 4C
in the light. 1cm crown sections and green leaf tissue were
separately harvested after 1, 3, and 6 hours of low
temperature exposure. The last 6 populations: After 7 days
of growth at 20C, wheat plants were transferred to 4C in
the dark. 1cm crown sections and green leaf tissue were
separately harvested after 1, 3, and 6 hours of low

temperature exposure. First strand synthesis in this
library was done in the presence of methylated dCTP
thereby protecting from internal cleavage with NotI. In
addition, this library used a primer for second strand
synthesis that annealed to an artificial sequence (RNA
oligo) added before first strand synthesis. Therefore when
sequences from EST generated from this library will be
masked for vector and adaptor sequences, an additional
masking step will have to be included to mask this RNA
oligo that is common to all clones (sequence
CGACTGGAGCAGGACACTGACATGACTGAAGAGTAGAAA)."

ORIGIN

Query Match 3.1%; Score 136.8; DB 7; Length 1083;
Best Local Similarity 52.8%; Pred. No. 1.9e-15;
Matches 391; Conservative 0; Mismatches 33; Indels 12; Gaps 4;
QY 3032 TCGTTCCCGCCTGTGGCATCGACGTACTGCGCGCGTGTATCCACCGTGTGATCGTCGCGAGGCCA 3091
Db 943 TTGTGCCCCCGTTGAGGCCCCAGACACTGGGCCCTGACCCACCCCGCGGTCTCTTGCAGA 884
QY 3092 GAAAGGCCACACAGTCGCGCGATGTCTGCGGCTCTGCCGATGCGGTTGAACACGAGTTGG 3151
Db 883 GGAACCCCAACGACCGCGCGATGTCTCGGCATGCCGAGCGCGCCACCGTCGCGCTCG 824
QY 3152 CGGCCAGTGGCGCGCGCGCTCGGGGTCTGCGCGCGCGCGTGTCTATGTCGCTCTCCA 3211
Db 823 CGATGTACCGCGCGCGCTCTCTCATCCGCTCTTCCCGTTG--TAGTACATCGGGGTGGCGG 766
QY 3212 CGAAACCCCGCGCGCCACCGGTTGACCGTGTATCCCGCGTTCCCGCGATTTCCCTGGCCAGGG 3271
Db 765 TGGACCCCGCGCGCCACCGGTTGCGGCTGTATCCCGTGTACCCCGAGCTCTCTTGGCCAGA 706
QY 3272 CGAGCGTGAAGTGTCCACCGCACCTTGGTTCATCGCTATCCGATGATCGAGTATCGGGGAACG 3331
Db 705 TCCTGTAATGCTCTCCACCGCGCGCTTGTCTACCGGTACGCGCGGTAGCGCGCGCGCA 646
QY 3332 CGCGCGGCTCGCGCGACAGAGATTTGATAT-----CGCGCGCGCTCGCGCAGTC 3385
Db 645 GCGACCCCGACCCCGCGACGAGAGAGGTGACATGCGCGCGCGCGCTCGCGGTGAGCC 586
QY 3386 GTTTCAGTCCGTGCTGGACAGCGGTGCGCGGACAGCGGTGCGCGGACCGAGTCCAGTCCGT 3445
Db 585 GTTTCGCGCGCTCCCGGAGGACAGAGAGTGTCCCGCGGTTGACACCGGAGCGCGCT 526
QY 3446 CGAAGACCTCTCTCGTGTGATTCCTCGTATCCGTTCGCGAGCGCTGAGCGCGCGGTGTTCA 3505
Db 525 CGTACGTGCGCTCGCTCGTCTCCGCGCGCGGTGTACGATGTCCAGACCGCGCGCTG 466
QY 3506 CCAGGATGTGAACCTCGGCGCGGAC--TCGAACTGCGCCATCCCGCGCTCGAACCGCGCG 3564
Db 465 TCGTGACGAGCACGTGTCAGCTCCCATCCGAAAGCGCGCGCGCTCGAACAGCGCC 406
QY 3565 TAGAGCGCGCGCGCTCACCCACAGCGCGAGTTGCGCGCGGATGCGCAACGCTGTCCG 3624
Db 405 CGCACATGCGCGCGCTCGACAGCTCCG--CTCCACCGGATGATGATGACCGTCCCG 349
QY 3625 CGCGTGTCTCGGATGTCTCGACGCTCTCTCGCGCGCGCGCTCGCTGTGCGGTAGTGG 3684
Db 348 TGGGCGCGGTTGATTCGCGCCACGACCTCTTGGCGCGCGCGCGCGCGCTAGGCC 289
QY 3685 ATGCGCACAGCGCGCGCTCGCGGCGACCGCGAGGCGGATACCGCGTTCGATGCCCGCG 3744
Db 288 ACCGCGCACGCGCGCGCGCGAGGCTGTCTAGGTGTTTGTGACACGCGCGAGCGCGGTGCGCGCG 229
QY 3745 CTTTCGCGCGTCCACGAGGC 3764
Db 228 GCACGCGCGTCAAGATGCG 209

RESULT 7
CL948439/c
LOCUS
CL948439 1257 bp DNA linear GSS 21-SEP-2004

```
DEFINITION OsIFS005157 Oryza sativa Express Library Oryza sativa (indica
cultiivar-group) genomic, genomic survey sequence.
ACCESSION CL948439
VERSION CL948439.1 GI:52360448
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 1257)
AUTHORS Ma L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G.K.S., Deng, X.W. and Wang, J.
TITLE An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
source
1..1257
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/notes="Oryza sativa exon trapped genomic sequences"
ORIGIN
Query Match 3.1%; Score 136.2; DB 10; Length 1257;
Best Local Similarity 53.1%; Pred. No. 2.4e-15;
Matches 313; Conservative 0; Mismatches 273; Indels 3; Gaps 1;
QY 3193 GCGTTTCATGTCGCTCTCCACGAAACCGCGGCCACCGGTGTGACCGTGATCCCGCGTTCC 3252
DB 590 GCGTGGATCTGGTATCGATCAGCGCGGCGCGACCGGTTTCACGCGGATGCCCTCCGCT 531
QY 3253 CCCAGTTCCCTGGCCAGGCGAGGTGAGCGGTGTCACCGCACCCCTTGGTCAATCGCGTAT 3312
DB 530 GCCACTTCGCGCGGAGGCCACCGTGAGCAGCTCGATGGCGGCCCTTGGCTGCGGCATAG 471
QY 3313 CCGATGGACTCGGGAAACGCGC--GCCGGGTGCGGCGAGCAGCAGATGTTGATGATCGC 3369
DB 470 TCCACGTACTGGTGGGCGGCGCCACCGCGGCGCGCTGGAAACATTGACGATGCTG 411
QY 3370 CGCGCGTCGCGCAGTCGTTTCAGTCCGTGCTGGACACGAGAAACAGCGGTGCGCGACGTTG 3429
DB 410 CGCCCTCGCGCGCGCTGCGAGGCTCATGCGCCGCACGGCTCGCGCGCAGTAGAAG 351
QY 3430 ACGGCGACAGTCGGTGAAGACCTTCCTCGGTGACTTCGGTATCCGTCGCGAGCGCGTG 3489
DB 350 CTGCCACGACATTGATGTGGAACATGCGCGCAGCGCTGGCGGCTCATCTCGGACAGC 291
QY 3490 ACGCCCGGTTGTTTACACAGATGTGCAACTCGGGCGGCATCTCCGAACTCGGCCCATCCG 3549
DB 290 CGTGAGGACAGTCCTACACCGCGCATTTGTTTACACGAGGCGGACAGCGCGGCCCATTCG 231
QY 3550 GCGTGAACCGCGCTAGAGCGCGCGCGCTCACCCACAGCAGCGGAGTTTCGGCCCGGATG 3609
DB 230 GTGTCCACCGCTCGAAACAGCGCCACGACCTCGGACTGTGCGGCACGTCGCGCTGGATG 171
QY 3610 GCCAAGCGCTGTTCGCGCGTGTCTCGGATGTTCTCGACGTTCTCTCGCGCGCGCGCTCG 3669
DB 170 GCCACGGCTTCCCGCCCTTCGCGCCCGATGCGATCGAACACGCGCGCGCGCGCGCG 111
QY 3670 CTGCTGCCGTAGTGACTGCGACAGAGCGCCCGTTCGCGGCCAGCGCGAGGCGGATACCG 3729
```

```
110 TCGCGGCATAGTTTCACTCCACGGCCAGCGCGCGCCAGCGCGCGTGGCG 51
QY 3730 CGTCCGATGCCCGGCTTCCCGCGGTACAGGGCGGTCTTGCCCTCA 3778
DB 50 GCGCGGATGCCCGGCTCCCTCCCGTCCCAAAACCACTTTTGTGCA 2
RESULT 8
CN823495 843 bp mRNA linear EST 02-JUN-2004
LOCUS Oa脾bn_05P05_M13reverse Sheep spleen\brain pSport1 library Ovis
DEFINITION aries cDNA clone Oa脾bn_05P05 5', mRNA sequence.
ACCESSION CN823495
VERSION CN823495.1 GI:47951564
KEYWORDS EST.
SOURCE Ovis aries (sheep)
ORGANISM Ovis aries
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Caprinae; Ovis.
REFERENCE 1 (bases 1 to 843)
AUTHORS Gossner, A. and Hopkins, J.
TITLE Ovine spleen\brain cDNA library
JOURNAL Unpublished (2004)
COMMENT Contact: J Hopkins
Veterinary Biomedical Sciences
University of Edinburgh
Summerhall Square, Edinburgh, EH9 1QH.
Email: j.hopkins@ed.ac.uk
Plate: 05 row: P column: 05
Seq primer: M13reverse
High quality sequence start: 16
High quality sequence stop: 547.
FEATURES
source
1..843
/organism="Ovis aries"
/mol_type="mRNA"
/db_xref="taxon:9940"
/clone_lib="Oa脾bn_05P05"
/clone_lib="Sheep spleen\brain pSport1 library"
/notes="Vector: pSport1"
ORIGIN
Query Match 3.1%; Score 135.6; DB 7; Length 843;
Best Local Similarity 53.7%; Pred. No. 3.2e-15;
Matches 388; Conservative 0; Mismatches 305; Indels 30; Gaps 4;
QY 3060 GCCGTGATCCACCGTGAAGTTCGAGGCCAGAAAGCCACCGTCGCGGATGTCGTC 3119
DB 109 GCCGTGATATAGTCGGCTCAGGGCTTACCGAGTAGGCGAGAACTGGCGATCTTTC 168
QY 3120 GGGTCTGCGGATGCGGTTGAACACGAGAGTTGGCGCCAGTGGCGCGCGCTCGGGGT 3179
DB 169 CACATGGCCATACCGGCCACGGCCATGAAGCCGATCAG-----GCTCTCGCGAA 219
QY 3180 CTGCGCGCGCGTGGTTCATGTCCTCCACGAAACCGCGGCCACCGGTTGACCGT 3239
DB 220 ATCACCACCTTCGCGGGTTTCATGTCGGTGTCCACCGGCTTCGCTGACGTTATTGATGT 279
QY 3240 GATCCCGGTTTCCCGCAGTTCCTTGGCCAGGCGGAGCGTGTGAGCGTGTCCACCGCACCCCTT 3299
DB 280 GATGCGACGTGGCCCGCAGTTCGCGGCCAGGCTTTGTGTCAGGCCGACGAGCGCGACTT 339
QY 3300 GGTTCATCGCGTATCCGATGGAATCGGGGAACG---CGCGCGGGTTCGCGGAGACGAGAT 3356
DB 340 GCTCATGCGTACGCGGCCGCCACCGGCAACCGCATGCGATCGGCAATTTGGTGTCTCCGAT 399
QY 3357 GTTATGATCGCGCGCGCTCGCGAGTCGTTTTCAGTCCGTCGTGGACGAGAAACAGCGG 3416
DB 400 GTTATGATGCGCGCACCTTCAACCATGTGCGCGCGCTTCTCTGGTGGGCGACAAACAC 459
QY 3417 TGCCCGGACGTTGACGCGGCGACGAGTCGCTCGGTCCGAAGACCTCTCTCGGTGACTTTCGATCCG 3476
```


Db	460	GCTGCGCAGCTTGTATGGCGAGGTTACGGTTCGAAGTCTTCCAGGCTGAAATCTTCCAGCGG	519
Qy	3477	TCCCGAGCGCTGACGCCCGGTTGTTCCACAGGATGTGCAACTCGGCGGCACTCCGAA	3536
Db	520	GCCGAGCGGCAGCAGCGCGGTTGTTTCCACAGATATCCA-----GGCGCCGAA	570
Qy	3537	CTCGGCCATCCCGCGTCAACCGCGGTAGAGCGCGCGGTCAACCCAGCAGCGCGAG	3596
Db	571	GCGTTGAGCGGTGCGGTTGACGCGCTCGGATGCGCGCGCAT-----CGGCGCT	621
Qy	3597	TTGCGCCGGATGGCCAAAGCGCTGTCGCGCGTGTCCGAGTGTCTCGAGGCTCTCTCG	3656
Db	622	GTGCGGATGATTCGCGCAGGCGCTTTCGCCGCTTCGCTGATCAGCTGTTCTGCAATGCTTC	681
Qy	3657	CGCGCGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	3716
Db	682	GSCCTTGCGCGCGGAGCTGACATAGTGAAGCGGAGCGCTGCGCTGNGCGCGCAGCGG	741
Qy	3717	CAGGCGGATACCGCTCGATGCCCGCGGCTTCCCGGCTCACAGGCGGCTCTTGGCCCTC	3776
Db	742	CTTGACGATGGCGCGCGGATACCGCGGACCGCGCTGATCAGGCGACTTTTCCGCGTG	801
Qy	3777	CAG 3779	
Db	802	AAG 804	
RESULT 9			
CN145046/c			
LOCUS			
DEFINITION			
WOUND1_26_B03_g1_A002 Wounded leaves Sorghum bicolor cDNA clone			
WOUND1_26_B03_A002 5', mRNA sequence.			
CN145046			
EST.			
CN145046.1 GI:45983920			
Sorghum bicolor (sorghum)			
Sorghum bicolor			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD			
clade; Panicoideae; Andropogoneae; Sorghum.			
1 (bases 1 to 800)			
Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C.,			
Sun,F., Sullivan,R., Brady,J., Eastman,A., Miller,V., Gonzalez,M.,			
Anfuso,C., Chhabra,D., Johnson,H., Kamran,D. and Pratt,L.H.			
A Sorghum EST database: mechanically damaged and methyl			
jasmonate-treated leaves			
Unpublished (2003)			
Other_ESTs: WOUND1_26_B03.b2_A002			
Contact: Cordonnier-Pratt MM			
Laboratory for Genomics and Bioinformatics			
The University of Georgia, Department of Plant Biology			
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA			
Tel: 706 542 1860			
Fax: 706 583 0210			
Email: mmprratt@uga.edu			
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in			
the Human Genome Center, University of Tokyo Institute of Medical			
Science, plant material and RNA prepared at Texas A & M University;			
sequencing done in the Laboratory for Genomics and Bioinformatics,			
University of Georgia. Sequence ends have been trimmed to exclude			
vector and regions below Phred quality 16. Three-prime sequences			
are presented as their reverse complement and have been trimmed to			
exclude polyA.			
Seq primer: Sug5 (CTTCTGCTCTAAAGCTGCG).			
Location/Qualifiers			
1. .800			
/organism="Sorghum bicolor"			
/mol_type="mRNA"			
/cultivar="BTx623"			
/db_xref="taxon:4558"			
/clone="WOUND1_26_B03_A002"			
/lab_host="DH10B-T1 phage-resistant E. coli"			
/clone_lib="Wounded leaves"			
FEATURES			
source			
RESULT 10			
CN2544259			
LOCUS			
DEFINITION			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY-2005			
SRAA-aad54h02.g1 Strongyloides ratti whole genome shotgun library			
(SRAAGSS 004) Strongyloides ratti genomic, genomic survey sequence.			
CN2544259			
CN2544259.1 GI:64674801			
GSS.			
Strongyloides ratti			
Strongyloides ratti			
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
Panagrolaimoidea; Strongyloidea; Strongyloidea.			
1 (bases 1 to 825)			
REFERENCE			
AUTHORS			
Mitrevva,M., McCarter,J.P., Thompson,F., Viney,M., Pape,D.,			
GSS 13-MAY			

Ritter,E., Martin,J., Wylie,T., Dante,M., Waterston,R.H., Clifton,S.W. and Wilson,R.
Genome Survey sequences from the rat parasitic nematode Strongyloides ratti
Unpublished (2005)
Contact: Mitreva M
Washington University in St. Louis
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: nematode@wustl.edu
Genomic DNA was provided by Fiona Thompson (F.Thompson@bristol.ac.uk) and Mark Viney (Mark.Viney@bristol.ac.uk) at the University of Bristol, Bristol, UK.
Class: shotgun.
Location/Qualifiers
1. .825
/organism="Strongyloides ratti"
/mol_type="genomic DNA"
/strain="isofemale line ED321 heterogonic"
/db_xref="taxon:34506"
/dev_stage="infective larval stage (il3)"
/lab_host="GS10"
/clone_lib="Strongyloides ratti whole genome shotgun library (SPRAGSS 004)"
/note="Vector: pORW13; Site 1: BstX1; Site 2: BstX1; Strongyloides ratti genomic DNA was randomly sheared, end-repaired and size fractionated to enrich for 2-4 kb fragments. Genomic DNA was provided by Fiona Thompson (F.Thompson@bristol.ac.uk) and Mark Viney (Mark.Viney@bristol.ac.uk) at the University of Bristol, Bristol, UK. Sequencing by Washington University Genome Sequencing Center, St. Louis, MO."

ORIGIN

Query Match 3.1%; Score 134.2; DB 10; Length 825;
Best Local Similarity 56.0%; Pred. No. 5.8e-15;
Matches 275; Conservative 0; Mismatches 213; Indels 3; Gaps 1;
Qy 3037 CCGCCTGTGGCATCGACGTACTGCGCGGTGATCCACCGTGTGATCGTGGAGCCAGAAAG 3096
Db 241 CCGCGTTGGCGCGCAGGATCTGGCGGTGATCCCGCAGCTGCGGCGCGCGCAGGAAG 300
Qy 3097 GCCACCACTGCGGATGTCTGGGCTGCGCGATGCGGTTGAACACGAGTTGCGGCC 3156
Db 301 GAAACCAACGAGGCGATGTCTTCGCGCTGACCCAGGCGCTGCAGCGCGCGCGCTTCGCA 360
Qy 3157 AGTGGCGGCGCGCTCGGGGTTCTGCGCGCGCGTGTTCATGTCTCGTCTCCACGAA 3216
Db 361 AAGTTCGGATCTGCTCTCGGTCTTGGCGCGAGGAAGCG---CCGTGGCGATCGGG 417
Qy 3217 CCGCGCGCACCGCTTGACCGTGTATCCCGCTTCCCGCAGTTGCTGGCAGGCGCAGC 3276
Db 418 CCGCGCGCACCGCTTGACCGTGTATGTCGGCGCGCGCAGTCTCTGCGAGACGTGC 477
Qy 3277 GTGAGCGTGTCCACCGCACCTTGTGTCATCGCGTATCGGATGATCTCGGGAAACCGCGCC 3336
Db 478 GTGAAGGCTTCACCGCGCGCTTGGTGGGTTGTAGACCGCATAGCCCGCGCAGGTTTCAGC 537
Qy 3337 CGGTCGGCGCAGAGATGTGATGATCGCCCGCGTTCGGCGAGTCTTCAGTCCG 3396
Db 538 GCGAGCACCGTGTCTCGAATAATGACGATGCGCCCGCCCTCTGTGAGACGCCCGCGGCC 597
Qy 3397 TGTGAGCAGGAAACAGCGGTGTCGGGAGCGTGTGACGGGACCGACGTCGTCGAAGACCTCC 3456
Db 598 TCGCGACGGTGTGAAGTGTGCGGACAGTGTATGTTCGAAGTCTGTGCTGTAGAGGTG 657
Qy 3457 TCGGTGACTTCGGTATCCGTCGCCGAGCGCTGACGCGCGGTGTTTACACGAGTATCG 3516
Db 658 TCGGTGTGCTCGGCGCAGCGGAGGTTCTTGAACACGCGCGCGGTATTATTCACGACGCTCG 717

Qy 3517 AACTCGGCGCG 3527
Db 718 ACCTTGCCAG 728
RESULT 11
DR887252
LOCUS
DEFINITION
DR887252 890 bp mRNA linear EST 01-AUG-2005
IMAGE:7796552 5', mRNA_sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Xenopus tropicalis (western clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE
1 (bases 1 to 890)
Richardson,P., Lucas,S., Rokhsar,D., Detter,J.C., Ng,D.C., Brokstein,P. and Lindquist,E.A.
DOE Joint Genome Institute Xenopus tropicalis EST project
Unpublished (2004)
Other ESTs: JGI_CUNC695.rev
Contact: Lindquist,E.A., Richardson,P.
DOE Joint Genome Institute
2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Tel: 925 296 5600
Fax: 925 296 5710
Email: cdna@jgi-psf.org
Tissue Procurement: Dan Buchholz (Yun-Bo Shi Laboratory, NIH)
cDNA Library Preparation: DOE Joint Genome Institute:
http://www.jgi.doe.gov
DNA Sequencing: DOE Joint Genome Institute: http://www.jgi.doe.gov
Clone Distribution: I.M.A.G.E. Consortium/LINL:
http://image.llnl.gov
Naming Conventions: EST name is generated by the concatenation of the JGI Clone Id and the direction of sequencing. The suffix '.fwd' indicates a forward sequencing read of the insert. It does not necessarily reflect the orientation of the insert.
Plate: CUNC 0005 row: n column: 6
High quality sequence stop: 852.
Location/Qualifiers
1. .890
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:7796552"
/tissue_type="whole embryo"
/dev_stage="Metamorphic (st.64)"
/lab_host="ElectromAX DH10B"
/clone_lib="NIH_XGC_tropMet4"
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This library was made from dr primed cDNA and cloned into Invitrogen pCMVSPORT6 vector. The work was done at DOE Joint Genome Institute. Poly A RNA were primed with 5' GACTAGTCTAGATCGCGAG CGCGCGCTTTTTTT 3'. CDNA were ligated to SalI adapter (5' TCGACCCACGGTCCG and 5'CGACGCGTGG), digested with NotI, size fractionated in 1.1% agarose gel electrophoresis and ligated into NotI and SalI digested pSPORT1 vector."
ORIGIN
Query Match 3.1%; Score 134.2; DB 8; Length 890;
Best Local Similarity 53.7%; Pred. No. 5.8e-15;
Matches 386; Conservative 0; Mismatches 303; Indels 30; Gaps 4;
Qy 3060 GCCGTGTATCCACCGTGTAGTGTGTCGAGGCCAGAAAGCCACCGTCGGCGATCTCGTC 3119
Db 94 CCGGTGTATATAGCGCGCTCAGGCGCTACCGAGTAGCGGACGAACTCGGATCTCTC 153
Qy 3120 CGGTCTGCCGATGCGGTTGAACACGAGTTGGCGCCAGTGGCGCGCGCTCGGGGT 3179

Db 154 CACATGGCCATACCGGCCACCGCCCATGAAGCGGATCAG-----GCTCTCGCGGNA 204
Qy 3180 CTGCGCGCCCGTGGTTCATGTCGTCTCCAGAAACCGCGCCACCGCGTTGACCGT 3239
Db 205 ATCACCACCTTGGCGGGTTCATGTCGTGTCCACCGGCCCTGGCTGCACGTTATTGATGT 264
Qy 3240 GATCCCGCGTTCCTCCAGTGTGCTGCGCAGGCGGAGCGTGAGGGTGTCCACGCGACCCCTT 3299
Db 265 GATGCCACGTGGCCCGCCAGGTGCGCGGCCACGGCTTTGGTTCAGCGCCACCGCGCCGACTT 324
Qy 3300 GGTCTATCGCGTATCCGATGACTCGCGGGAACG---CGCGCGGGTTCGCGGCAGACGAGAT 3356
Db 325 GCTCATGGCGTACGGGCCCGCCACCGGCACCGCATCGCATCGCATGCTGCTGCTGCGGAT 384
Qy 3357 GTTGTATGATCCCGCGCCGCTCGCGCATGCTGTTTCAGTTCGCTGTCGACGAGCAAGCGG 3416
Db 385 GTTGTATGATCGCGCCACCTTCACCATGTCGCGCGGGCTTCCTGGGTGGCGCAACAC 444
Qy 3417 TCGCGGAGCTTGACCGGCACACGTCGTCGAGACCTCTCGTGCATCTCGTGTATCGG 3476
Db 445 GCTGCGCACGTTGATGGCGAGGTACGGTTCGAAGTCTTCAGGCTGAATCTTCCAGCGG 504
Qy 3477 TCCGAGCGCGTGACGCGCGCTGTTTTCACGAGGATGTCGAATCGCGGCGCACTCCGAA 3536
Db 505 GCGGACGCGCAGCACCGCGGCTGTTTTCACGAGATATCA-----GGCGCCGNA 555
Qy 3537 CTCGCCATTCGCGGCTCGAAGCGCGGTAGAGCGCGCGGTACACCGACGCGCG 3596
Db 556 GCGTTTCAGCGGTGGCTTACGCGCGCTGCGGATGGCGCGCAT-----CGCGCT 606
Qy 3597 TTCGCGCGGATGGCCACGCTGTCGCGCGCTGTCGCGGATGTCGCGGATCTCTCG 3656
Db 607 GTCGGATGATCGCCAGGCGCTTTCGCGCTGCTGCTGATCAGCGTCTTCGAATGCTTC 666
Qy 3657 CGCGCGCGCTCGTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3716
Db 667 GCGCTTGGCGCGCGAGCTGACATAGTGAAGCGGACGCGCTGCGCTTGGCGCGCGAGG 726
Qy 3717 CAGGGCGATACCGCGTTCGATGCGCGGCTTCCCGCGGTACACGAGCGGCTTTCGCCCT 3775
Db 727 CTTGAGATGGCGCGCGCGATACCGCGGACCGCGCTTGGATCAGCGACTTTCGCGCT 785

RESULT 12

CN130030/c
LOCUS
DEFINITION
RHOH1_38_H01.g1_A002 Acid- and alkaline-treated roots Sorghum
bicolor cDNA clone RHOH1_38_H01_A002 5', mRNA sequence.

Accession
CN130030

Version
CN130030.1

Keywords
GI:45958210

Source
EST.

Organism
Sorghum bicolor (sorghum)

REFERENCE
Sorghum bicolor (sorghum)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 764)

Cordonnier-Pratt M.-M., Suzuki Y., Sugano S., Klein R.R., Liang C.,
Sun F., Sullivan R., Lim S., Eastman A. and Pratt L.H.

An EST database from Sorghum: acid- and alkaline-treated roots

Unpublished (2003)

Contact: Cordonnier-Pratt M.

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in

the Human Genome Center, University of Tokyo Institute of Medical

Science; plant material and RNA prepared at Texas A & M University;

sequencing done in the Laboratory for Genomics and Bioinformatics,

University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below Phred quality 16. Three-prime sequences
are presented as their reverse complement and have been trimmed to
exclude polyA.

Seq primer: Sug5 (CTTCTGCTCTAAAGCTGG).

Location/Qualifiers

1..764

/organism="Sorghum bicolor"

/mol_type="mRNA"

/cultivar="BTx623"

/db_xref="taxon:4558"

/clone="RHOH1_38_H01_A002"

/lab_host="DH10B-T1 phage-resistant E. coli"

/clone_lib="Acid- and alkaline-treated roots"

/note="Organ: Root; Vector: pME18S-FL3; Site: 1; XhoI;

Site 2: XhoI; The library was prepared from polyA+ RNA

from 8-day-old roots harvested from BTx623 sorghum

seedlings grown in hydroponic culture. HCl was added to a

pH of 3.0 to some seedlings, KOH to a pH of 9.0 for

others. Roots were harvested 3, 12 and 27 hr after

addition and pooled for RNA preparation. Double-stranded

cDNA was cloned unidirectionally into different DraIII

sites of the pME18S-FL3 vector (5-prime DraIII site is

CACGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises

the cDNA insert."

ORIGIN

Query Match 3.1%; Score 133.8; DB 7; Length 764;
Best Local Similarity 54.5%; Pred. No. 6.9e-15;
Matches 320; Conservative 0; Mismatches 252; Indels 15; Gaps 2;

Qy 3202 TCGGTCTCACAAACCGCGGCCACCGGTGACCGGTATCCCGTTCCTCCCACTTCG 3261
Db 696 TCGGTGCGCACGGGCGCGCGGACACGTTTCGAGTCCTCCCTTCGCGCGCACCTCC 637
Qy 3262 CTGGCCAGCGCGAGCGTGTGTCACCGCACCTTCGTATCCGCTATCCGATGAGAC 3321
Db 636 TTGGCTAGGATCTTGTATCTCCCTCCAGCGGCGTGTGTCGCGGTGTACGCGCGTAT 577
Qy 3322 TCGGGGAACGCGCGCGGTTCGCGCAGACGAGATTTGATGATCCGCGCGC-----G 3375
Db 576 CCGCGCAGGAGCGTTCGCGACGATGACGACGAGAACGTCACGATGCGCGCGCGCTGTTG 517
Qy 3376 TCGCGCAGTCTGTTTCAGTCTGTCGACGAGAACGAGGAGGTCGCGGAGCTTACGCGCG 3435
Db 516 GCTGGGACTTCGTTTCGCGCCTCACGSCACACGAGAACGTCGCGGCGACGTTTCACGCG 457
Qy 3436 ACCAGTTCGTCGAAGACCTTCCTCGGTGACTTCGCTGATCCGTCGCGCGCGCTGACGCGC 3495
Db 456 AACATGGCGTGAAGTCTTCGACGCGCGTGTTCGCGAGCGCGCGGTACTTGTGCTGAGG 397
Qy 3496 GCGTTGTTTTCACAGGATGTGAACTTCGGCGGCGCACTTCGGAATTCGCCCATTCGCGGCGT 3555
Db 396 ACGCTTCACAGCAGACGACGATGTTCGCGCGGCGGACCGCAACGCTCTTCGCGCGCGT 337
Qy 3556 AACGCGCGTTCAGAGCGCGCGCGGTTCACCGACGCGCGGATTCGCGCGCGGATTCGCGC 3615
Db 336 AAGAGCGCGCAGCAGCGCGTTCGCGGTCTGACAG-----TCCGCGGCGAGCGCGACG 286
Qy 3616 GCGTGTTCGCGCGTCTTCGCGGATGTCGACGCGTCTTCGCGCGCGCGCGCTCGCTGCTG 3675
Db 285 GCGTGTTCGCGCGGAGCGGAGCTTCGCGACGAGTCTGCTGCTTCCTTCGCGGGAATTGGAC 226
Qy 3676 CCGTAGTGACTTCGCAACGAGCGCGCGTTCGCGCGCGAGCGCGGAGGCGGATACCGCGTCCG 3735
Db 225 GCGTAGTTTCACCAAGCAGCGCGCGCGGCGGCGGAGGTTGGAGGACACCTTCGCGCGCG 166
Qy 3736 ATGCCCGCGGCTTCCCGCGGTTCACGAGCGCGGTTCGCGCGCGGCTTCGCGCTCCAGCGG 3782
Db 165 ATGCCCGCGGAGCGCGCGGTTCGAGCGCGGCGCGCGCGCGCTTCGAGCGCG 119

RESULT 13

```
CO202814      780 bp      mRNA      linear      EST 21-JUN-2004
LOCUS      Oa_splbn_14N14_M13_Reverse_Sheep_spleen\brain pSpor1 library Ovis
DEFINITION      aries cDNA clone Oa_splbn_14N14 5', mRNA sequence.
ACCESSION      CO202814
VERSION      CO202814.1  GI:49013989
KEYWORDS      EST.
SOURCE      Ovis aries (sheep)
ORGANISM      Ovis aries
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactylia; Ruminantia;
Pecora; Bovidae; Caprinae; Ovis.
1 (bases 1 to 780)
Gosner,A. and Hopkins,J.
Ovine spleen\brain cDNA library
Unpublished (2004)
Contact: J Hopkins
Veterinary Biomedical Sciences
University of Edinburgh
Summerhall Square, Edinburgh, EH9 1QH.
Email: j.hopkins@ed.ac.uk
Plate: 14  row: N  column: 14
Seq primer: M13 Reverse
High quality sequence start: 15
High quality sequence stop: 546.
FEATURES             source
     source
     1..780
     /organism="Ovis aries"
     /mol_type="mRNA"
     /db_xref="taxon:9940"
     /clone="Oa_splbn_14N14"
     /clone_lib="Sheep spleen\brain pSpor1 library"
     /note="Vector: pSpor1"

ORIGIN
Query Match      3.0%; Score 132; DB 7; Length 780;
Best Local Similarity 54.0%; Pred. No. 1.5e-14;
Matches 376; Conservative 0; Mismatches 290; Indels 30; Gaps 4;

QY 3060 GCCGTGATCCACCGTAGTCGTCCGAGGCCGAAAGCCACACGTCGGGATGTCGTC 3119
DB 109 GCCGTGATATAGCCGGCTCAGGCGCTACAGGTAGGCGACGAACTGGCGATCTCTTC 168
QY 3120 GGGTCGCGATGCGGTGAACACCGAGTTGGCGCCAGTGC CGCGCGCTCGGGGT 3179
DB 169 CACATGGCCATACCGGCCCGCCAGCCATGAAGCCGATCAG-----GCTCTCGCGAA 219
QY 3180 CTGCGCGCGCGGTGGTTCATGTCGTCTCCACGAAACCGCGCGCCACCGCTTGACCGT 3239
DB 220 ATCACACTTTCGGGGTTCATGTCGTCTCCACCGGGCTTGCTGCACGTTATTGATGGT 279
QY 3240 GATCCCGCGTTCCTCCAGTTGCTCGGCCAGGCGAGCGTGAGCGTGTCCACGCAACCTT 3299
DB 280 GATGCCAGTGGCCCCAGGTGCGCGGCCAGGCTTTGGTCAGGCCGACGAGCGCGACTT 339
QY 3300 GGTATCGCGTATCCGATGACTTCGGGNAAG---CGCGCGGTTCGGCGCAGACGAGAT 3356
DB 340 GCTCATGCGTACGGCGCGCCACCGGCCAAACGGCATGCGATCGGCAATTGGTCTGCCGAT 399
QY 3357 GTTATGATTCGCCCGCGCTGCGCAGTCTGTTTCAGTCCGTCTCGACGAGAACAGCGG 3416
DB 400 GTTATGATGCGGCCACCTTTCACCATGTCGCGCGGGCTTCTGGGTGGCAGCAACAC 459
QY 3417 TGCCCGGAGTTGACGGCGACAGTCGTCGAGAACCTCTCTCGGTGATCTTCGATCCG 3476
DB 460 GCTGCGCAGTTGATGCGGAGGTACCGGTTCGAAGTCTTCAGGCTGAAATCTTCCAGCG 519
QY 3477 TCCGAGCGGTGACGCCCGCGTGTTCACGAGATGTGCAACTTCGGGGGCACTCCGNA 3536
DB 520 GCCGACGGCCACGACCGCGCGTGTGTTACAGGATATCCA-----GGCGCCCGAA 570
QY 3537 CTCGCCCATCCCGCGGTTCGAACCGCGGTAGAGCGCGCGCGGTTCACCCACGACCGCAG 3596
```

```
DB 571 GGCTTGACAGGTGGCGTTGACGGCGCTCGGATCGCGCGGCAT-----CGGCGCT 621
QY 3597 TTCCGCCCGGATGCCAACGCTGTCCGCGCTGTCCCGATGFTCTCGAGGTCCTCG 3656
DB 622 GTCGGCATGGATCGCCAGGGCTTTGCCGCCCTTCGCTGATCACGCTGTCTGCAATGCTTC 681
QY 3657 CGCGCGCGCTCGCTGCTGCCGTAGTGACTGCCACGAGCGCCCGCTCGCGCGCAGCG 3716
DB 682 GGCCTTGGCGCGGCTGATAGTAGTGAAGCGAGCGCTCGGCTTGGCGCGCAGCG 741
QY 3717 CAGGCGGATACCGGTCCGATGCCCGCGCTTCCGCC 3752
DB 742 CTTGACGATGGCGCGCGGATACCGCGGACCCGCC 777

RESULT 14
LOCUS      CC398497/c
DEFINITION      PUHLC28TD ZM_0.6_1.0 KB Zea mays genomic clone ZMMBTA476B07,
genomic survey sequence.
ACCESSION      CC398497
VERSION      CC398497.1  GI:30878587
KEYWORDS      GSS.
SOURCE      Zea mays
ORGANISM      Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 809)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PUHLC28TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES             source
     source
     1..809
     /organism="Zea mays"
     /mol_type="genomic DNA"
     /strain="B73"
     /db_xref="taxon:4577"
     /clone="ZMMBTA476B07"
     /clone_lib="ZM_0.6_1.0 KB"
     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
     Cot selected genomic DNA library"

ORIGIN
Query Match      3.0%; Score 132; DB 9; Length 809;
Best Local Similarity 52.9%; Pred. No. 1.5e-14;
Matches 307; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

QY 3202 TCCGTCTCCAGAAACCGCGGCCACCGCTGACCGGTATCCCGCTTCCCGCAAGTTGC 3261
DB 762 TCCGTGCGCACAGGCGCGCGCCACACGTTGGCGGTGATCCCTTTCGCGCACTCC 703
QY 3262 CTGGCAGGCGGAGCGGTGAGGCTGTCCACCGCACCTTGGTCTATCGGATCCGATGAC 3321
DB 702 TTGGCCAGGACCTCTGTCATCGCTTCCACAGCGGCTTGGACGCGAGCTAGCGCGCTAG 643
QY 3322 TCGGGGACGCGCGCGGTGCGGCGACAGAGATGTTGATGATCCGCGCGCTCGCGC 3381
DB 642 CCCGCGGACGACGCGCAACCGTCTGTGGACGAGAACGTCATGATCGCGCCACCGCTTTG 583
QY 3382 AGTCGTTTCAGTCCGTGCTGGAGCCAGGAACGCGGTGCCCGACGTTGACGCGCAACGAT 3441
```

```

Db      582 GGGGGGATCGCGCGCGGCGCTCCCGGACACACAGGAATCTCCGCGCGGTGTTACCGGG 523
QY      3442 CGGTGAAAGACCTCTCTCGGTGACTTTCGGTGATCCGTCCGAGCCGCTGACGCCCGCGTTG 3501
Db      522 AACGTGGCGTTCGAGTCTCGACGCGAGTTCGGCCAGCGCGGGTACTTGGCGTCCATG 463
QY      3502 TTCACAGAGATGTCGAACTTCGGGCGGCACTCGAACTCGCCCATCCGCGGTGGAACGCC 3561
Db      462 ATGCCCGCGCAGCGGACCACTACGTGGGCGGGGACCCGAAATGCCCTCTCCGCGCGG--- 406
QY      3562 GCGTAGAGCGCGCGCGCTCACCAAGCGCGAGTTCGGCCCGGATGCGCCAAAGCGCTGT 3621
Db      405 TGAAGAGCGCGCGACAGCGCGCGGTCGGAGAGCTTCGCGCGGCAAGGCCACAGCGCGG 346
QY      3622 CGCGCGCTCTCGGATGTTCTCGAGGTCTTCGCGCGCGCGCGCTTCGTGTGCGGTAG 3681
Db      345 AGCGCGCGGAGCGAGCTCCGCGACGAGCTCTCTCGCGCTTCGCGCGGTGCGGTAG 286
QY      3682 TGGACTGCCACAGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3741
Db      285 TTGACCAACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 226
QY      3742 CGCGCTTCCCGCGTACACAGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3781
Db      225 CGGAGACCAACGCGTACAGCGCGGATGCGCGCTTCGAGCG 186

RESULT 15
BZ537000/c
LOCUS   BZ537000 795 bp DNA linear GSS 16-DEC-2002
DEFINITION OGAGU627C ZM2_0.7_1.5_KB Zea mays genomic clone ZMMBMA0058K03,
genomic survey sequence.
ACCESSION BZ537000
VERSION   BZ537000.1 GI:27085218
KEYWORDS  GSS.
SOURCE    Zea mays
ORGANISM  Zea mays
REFERENCE 1 (bases 1 to 795)
AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
CONSORTIUM Consortium for Maize Genomics
JOURNAL   Unpublished (2002)
COMMENT   Contact: Cathy Whitelaw
TIGR      9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: methylation filtered.
FEATURES
    source
        1..795
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone="ZMMBMA0058K03"
            /clone_lib="ZM2_0.7_1.5_KB"
            /notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
            methylation filtered genomic DNA library"

ORIGIN
Query Match 3.0%; Score 131.6; DB 9; Length 795;
Best Local Similarity 53.5%; Pred. No. 1.8e-14;
Matches 387; Conservative 0; Mismatches 304; Indels 33; Gaps 4;

QY      3041 CTGTGGCATGACGACTGCGCGGTGATCCACGTCGTGCGGAGCGGAGGCGCA 3100
Db      698 CCGTTGGCGCGATGACCTGGCGGTGACCCACTCGCGCGCGTGGTGACAGGAGCGCG 639

```

```

QY      3101 CCACGTCGCGGATGTCGTCCGGTCTGCCGATCGGTTGAAACAGGAGTTGGCGGCGCATG 3160
Db      638 CCACGCGCGGATGTCACCGGGCTCGCGAGCCTCTGCATC---GGGTTGTCTCCACGG 582
QY      3161 CGCGCGCGCTTCGCGGGGTCTGCCGCGCGGTGCGTTTCATGTCCGTCTCCAGAAACCGG 3220
Db      581 CGTGTGCGACCATGCGCTCGCTCTTGGCCCTTGAAGAACATGTCCTGCGCGGTGGCCCGG 522
QY      3221 GCGCCACCGCGTTGACCGTGATCCCGCGTTCCCGAGTTGCTGCGCAGGGCGAGGTGA 3280
Db      521 GCGCGACGAGTTGGCGGTGACGCGGTGCCCGGAGCTCTTTGGCCATGCGTCCGACCA 462
QY      3281 GGGTGTCCACCGCACCTTGGTTCATCGGTATCCGATGACTCGGGAACGCGCGCGCG 3340
Db      461 GCGCTTCCACCGCCCTTGGACCGCGGTAGGCCCCGAAACCGCGCGGAGGCTCCCCA 402
QY      3341 TCGCGCGAGACGAGATGTTGATGATCGCCCCCGCGTC-----GCGCAGTCTGTTTCAGTC 3394
Db      401 CCACGGAAGAGTCAAGGCGAGATGCGCGCGCGCGCGCGCGCGCGCGCGGTGGCGG 342
QY      3395 CGTGTGGAACCAAGAACAGCGGTGCGCGGACGTTTGAAGCGGACAGTCTGCTCGAAGACT 3454
Db      341 CTTCCCGGAGGCAAGGAACGCGCGCAGGTTTCAAGCGGAGCAGCGCGTCCGAGGCC- 283
QY      3455 CCTCGGTGACTTTCGTTGATCCGTCCGAGCGCGTTCGAGCGCGCGGTGTTTCAACGAGATGT 3514
Db      282 -----TCGGTGGCGGTGCGCGGCGCACCGCGGGGTACGTGTGTCAGACACCCCGG 234
QY      3515 CGAACTCGGGCGGCACTCCGAACTCGCCCATCCCGCGTCGAAACGCGCGGTAGAGCGCGG 3574
Db      233 CCTGGCCACCAAGATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAGCGGAC 174
QY      3575 CGCGCTCACCCACAGCGCGAGTTCGGCGCGGATGCGCCAAAGCCTGTCGCGCGCTGCTCC 3634
Db      173 GCGCGCGCGCTCTGTCGAGAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 126
QY      3635 GATGTTCTCGAGCGGTCTCTCGCGCGCGCGCGCTCGTGTGCTGCGGTAGTGGACTGCCACGA 3694
Db      125 GGAGCGACCGCCCGCGCGCGCGCTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66
QY      3695 GCGCCCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3754
Db      65 TCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6
QY      3755 TCAC 3758
Db      5 TCAC 2

```

Search completed: January 18, 2006, 20:42:40
Job time : 15846 secs

This page blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2006, 09:42:01 ; Search time 2374 Seconds
(without alignments)
12200.823 Million cell updates/sec

Title: US-10-017-471B-19
Perfect score: 4346
Sequence: 1 gtcgacgacgcgcgcgggtt.....gcgtagagcgcgggtacc 4346

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4346	100.0	4346	6 ABN86379	Abn86379 ScbA, Scb
2	4346	100.0	4346	10 ADE14794	Adel14794 Streptomy
3	401	9.2	401	6 ABN86377	Abn86377 S. coelic
C 4	401	9.2	401	6 ABN86378	Abn86378 S. coelic
5	401	9.2	401	10 ADE14790	Adel14790 Streptomy
6	316.2	7.3	19016	8 AAD47222	Aad47222 Streptoco
7	274.2	6.3	88400	13 ADU47667	Adu47667 S. cyaneo
C 8	255.4	5.9	7584	2 AAX19362	Aax19362 Rhodococc
C 9	255.4	5.9	7600	3 AAAS1878	Aaas1878 Rhodococc
10	239	5.5	4721	14 AEA81294	Aea81294 Streptomy
C 11	238	5.5	936	13 ADW38421	Adw38421 DNA seque
12	229	5.3	699	6 AAL44719	Aal44719 S. virgini
C 13	219.8	5.1	14638	14 ACL64643	ACL64643 M. xanthu
C 14	216	5.0	744	14 ACL65960	ACL65960 M. xanthu
C 15	213	4.9	741	13 AD744601	Adt44601 Bacterial
C 16	210.6	4.8	762	13 ADS55812	AdS55812 Bacterial
C 17	210.6	4.8	762	13 ADS50097	AdS50097 Bacterial
C 18	206.6	4.8	750	13 ADS62269	AdS62269 Bacterial
C 19	206.6	4.8	750	13 ADS62467	AdS62467 Bacterial

C 20	206.6	4.8	750	13	ADS59685	AdS59685 Bacterial
C 21	206.6	4.8	750	13	ADS62885	AdS62885 Bacterial
22	196.6	4.5	24379	2	AAT93095	Aat93095 Streptomy
23	196.6	4.5	24379	2	AAV25925	Aav25925 Streptomy
24	194.2	4.5	687	14	AEA81295	Aea81295 Streptomy
C 25	185.4	4.3	732	13	ADS59899	AdS59899 Bacterial
C 26	184.6	4.2	912	11	ABD00980	Abd00980 Klebsiell
C 27	176.6	4.1	777	13	ADS59919	AdS59919 Bacterial
C 28	175.4	4.0	8169	2	AAV26609	Aav26609 Actinomad
C 29	174.6	4.0	762	13	ADS61934	AdS61934 Bacterial
C 30	172.8	4.0	765	13	ADS63580	AdS63580 Bacterial
C 31	172.8	4.0	765	13	ADS63953	AdS63953 Bacterial
C 32	172.6	4.0	747	13	ADS59965	AdS59965 Bacterial
C 33	171.4	3.9	747	13	ADS60268	AdS60268 Bacterial
C 34	169.2	3.9	753	13	ADT41762	Adt41762 Bacterial
C 35	167.6	3.9	732	13	ADS64319	AdS64319 Bacterial
C 36	164	3.8	726	13	ADS55246	AdS55246 Bacterial
C 37	162.8	3.7	720	13	ADS49731	AdS49731 Bacterial
C 38	162.2	3.7	786	8	ADA70196	Ada70196 Rice gene
C 39	162	3.7	765	13	ADS45896	AdS45896 Bacterial
C 40	161.4	3.7	906	11	ACH98155	Ach98155 Klebsiell
C 41	159	3.7	792	11	ABD12618	Abd12618 Pseudomon
C 42	159	3.7	936	11	ABD12561	Abd12561 Pseudomon
C 43	158	3.6	742	13	ADT45738	Adt45738 Bacterial
C 44	156.8	3.6	876	11	ABD13397	Abd13397 Pseudomon
C 45	156.4	3.6	114955	2	AAX53491	Aax53491 Human ade

ALIGNMENTS

RESULT 1
ABN86379
ID ABN86379 standard; DNA; 4346 BP.
XX
AC ABN86379;

DT 21-OCT-2002 (first entry)

DE ScbA, ScbR and ScbB encoding genes containing DNA sequence.

XX Antibiotic; bacterium; scbA; afsA; scbR; arpA; barA; actinorhodine; Act;
undecylprodigiosin; Red; gene; ds.

OS Streptomyces coelicolor.

PN CA2322241-Al.

PD 23-APR-2002.

PF 23-OCT-2000; 2000CA-02322241.

PR 23-OCT-2000; 2000CA-02322241.

XX (PLAN-) PLANT BIOSCIENCE LTD.

XX Takano E, Bibb M;

DR WPI: 2002-501089/54.

XX P-PSDB; AB880940, AB880941, AB880942.

PT Modifying antibiotic-producing Streptomyces, to increase, or alter timing
of, antibiotic production, by deleting the scbA or scbR genes.

XX Claim 19, 20; Fig 14; 64pp; English.

XX The invention provides a method for modifying an antibiotic-producing
strain of Streptomyces to increase production of antibiotics or to alter
the timing of antibiotic production. The modification is functional
deletion of the scbA gene of S. coelicolor, or its homologues, but is not
deletion of the afsA gene of S. griseus, or the modification is
functional deletion of the scbR gene of S. griseus, or its homologues,
but is not deletion of arpA of S. griseus nor barA of S. virginiae. The

method is particularly used for production of the antibiotics actinorhodine (Act) and undecylprodigiosin (Red). The present sequence represents the nucleic acid sequence of the region containing the S. coelicolor ScbA, ScbR and ScbB proteins

Sequence 4346 BP; 650 A; 1561 C; 1519 G; 616 T; 0 U; 0 Other;

Query Match 100.08; Score 4346; DB 6; Length 4346;
 Best Local Similarity 100.08; Pred. No. 0;
 Matches 4346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCACACGACGCGTCGGGTTCCGAGCCGACCGGTAATCTCGGTCGACCGGACCG 60
 Db 1 GTCGACGACGCGTCGGGTTCCGAGCCGACCGGTAATCTCGGTCGACCGGACCG 60

Qy 61 GGTCTGGCTCGATGACCGACCGATCGAGACGTCGGCTGGCGCTCTCGATAGTAC 120
 Db 61 GGTCTGGCTCGATGACCGACCGATCGAGACGTCGGCTGGCGCTCTCGATAGTAC 120

Qy 121 GGCCCGCGCGGACGACATCGAGCTCCATCTCCCACTGCGCCCGCGAAAGTGAGC 180
 Db 121 GGCCCGCGCGGACGACATCGAGCTCCATCTCCCACTGCGCCCGCGAAAGTGAGC 180

Qy 181 ACCGACCGCGAGCTGACGATCGAGGAGGCGACGTCGCGGACGATCAACCCCTGGCT 240
 Db 181 ACCGACCGCGAGCTGACGATCGAGGAGGCGACGTCGCGGACGATCAACCCCTGGCT 240

Qy 241 TCGGCGGAAGCTTCGCGTGGTCCGCGCCGACGATCGGAAAGCGCTGGCGGCGGCGC 300
 Db 241 TCGGCGGAAGCTTCGCGTGGTCCGCGCGCGACGATCGGAAAGCGCTGGCGGCGGCGC 300

Qy 301 CGTCACTCCGCGAGTCGCGACGACCTGCGCTGCGCCCGCGCCCGCTGCGCTGCGC 360
 Db 301 CGTCACTCCGCGAGTCGCGACGACCTGCGCTGCGCCCGCGCCCGCTGCGCTGCGC 360

Qy 361 GCTCTTCAGGAAAGCGTCGCGGACGCGGCTCGGATATCTTCGCGAGCATCTCTCGCGC 420
 Db 361 GCTCTTCAGGAAAGCGTCGCGGACGCGGCTCGGATATCTTCGCGAGCATCTCTCGCGC 420

Qy 421 CGCTTCAGCCCGAGCTCTCGAAACGACGATGACCGCCCGACGCTCGGAGAGCGCGC 480
 Db 421 CGCTTCAGCCCGAGCTCTCGAAACGACGATGACCGCCCGACGCTCGGAGAGCGCGC 480

Qy 481 CGTCTGCTGCTGCGGTCGATGCGCTCGGCGCGGACGCGGTCAGCCGCTGCTCGTG 540
 Db 481 CGTCTGCTGCTGCGGTCGATGCGCTCGGCGCGGACGCGGTCAGCCGCTGCTCGTG 540

Qy 541 GCCGGGCGGGTCAGTGCCCTGACCGCCACTTCGCGCATGTCGCGCGATCGACGAGC 600
 Db 541 GCCGGGCGGGTCAGTGCCCTGACCGCCACTTCGCGCATGTCGCGCGATCGACGAGC 600

Qy 601 AACCGGGACGTGCGGTACAGCGCGGACACACGCGTCGACCGGATGGCGGCGGCCA 660
 Db 601 AACCGGGACGTGCGGTACAGCGCGGACACACGCGTCGACCGGATGGCGGCGGCCA 660

Qy 661 GGACAGCTGTTGACATGAGAGTCTGCGCGCGGACGAGTCCAGTTCAGCCCGGACTC 720
 Db 661 GGACAGCTGTTGACATGAGAGTCTGCGCGCGGACGAGTCCAGTTCAGCCCGGACTC 720

Qy 721 CGGTACGCGCGCTGTTCTCGCGTCGCGCGGTCGATGAAGTCTGCGCGCGCGGTTTC 780
 Db 721 CGGTACGCGCGCTGTTCTCGCGTCGCGCGGTCGATGAAGTCTGCGCGCGCGGTTTC 780

Qy 781 CCCACCGCGAGCATGGACAGCTTACACAGTTCGCGGACGCGCGCTCGCGCGCGCGC 840
 Db 781 CCCACCGCGAGCATGGACAGCTTACACAGTTCGCGGACGCGCGCTCGCGCGCGCGC 840

Qy 841 CGCGAAACGCTCGTCTCGGCTCGGTCAGTCTGTTGTCGAGGAAACGCGCCCGCAC 900
 Db 841 CGCGAAACGCTCGTCTCGGCTCGGTCAGTCTGTTGTCGAGGAAACGCGCCCGCAC 900

Qy 901 CCCGTTGAGGCGCGGTCGAGCGCGGCGGTCGCGCTACTCGCCCGCGAGACCTCGAC 960
 Db 901 CCCGTTGAGGCGCGGTCGAGCGCGGCGGTCGCGCTACTCGCCCGCGAGACCTCGAC 960

Db 901 CCCGTTGAGGCGCGGTCGAGCGCGGTCGCGCTACTCGCCCGCGAGACCTCGAC 960
 Qy 961 GTTCGGGCGCGGTGACCGGTCAACCGGTCGCGCGCGCGGCGAGGACTCTGACGGGACCGGT 1020
 Db 961 GTTCGGGCGCGGTGACCGGTCAACCGGTCGCGCGCGCGGCGAGGACTCTGACGGGACCGGT 1020

Qy 1021 CCGGGCCAGCAGAGTGGCGACCTGACCGGCCAGACACACCGGTCAACCGGTCACAAAGAT 1080
 Db 1021 CCGGGCCAGCAGAGTGGCGACCTGACCGGCCAGACACACCGGTCAACCGGTCACAAAGAT 1080

Qy 1081 CACTCGGGGCTCTCTCGGCGAGCGAGGCGGCTCCGAAACATACATATGAGGGGA 1140
 Db 1081 CACTCGGGGCTCTCTCGGCGAGCGAGGCGGCTCCGAAACATACATATGAGGGGA 1140

Qy 1141 AGGCGAGATCTGCCCCGGGGCGGAACCGGCGATGTTGCGGCCCGCGGGCGGTCCTTC 1200
 Db 1141 AGGCGAGATCTGCCCCGGGGCGGAACCGGCGATGTTGCGGCCCGCGGGCGGTCCTTC 1200

Qy 1201 AGCGGAGAACCGGGGCGCGACAGCGTGTGTGAGAAAGACGAGGCTGCGCTCTGATGCC 1260
 Db 1201 AGCGGAGAACCGGGGCGCGACAGCGTGTGTGAGAAAGACGAGGCTGCGCTCTGATGCC 1260

Qy 1261 CGGTGACCCGCAACGCGTGTGTCAGCCCCCGCGCGCGCGCGCGCGCGCGCGTGGATTC 1320
 Db 1261 CGGTGACCCGCAACGCGTGTGTCAGCCCCCGCGCGCGCGCGCGCGCGTGGATTC 1320

Qy 1321 AGCAGGGCTGTGAACTCCCGCTACCGGACGAAACCGGGTCCCGGATCGACGGCAG 1380
 Db 1321 AGCAGGGCTGTGAACTCCCGCTACCGGACGAAACCGGGTCCCGGATCGACGGCAG 1380

Qy 1381 AGGCGCGCGACCGGTCAACGAGGCAACCGCTGCGCGCTCGAGCAGCAGCATGC 1440
 Db 1381 AGGCGCGCGACCGGTCAACGAGGCAACCGCTGCGCGCTCGAGCAGCAGCATGC 1440

Qy 1441 CCGGTACGTGCTGTTGCGGCGCTGGAAGAGGGTGGGTGATGCTGCTGTCACCCGCAATC 1500
 Db 1441 CCGGTACGTGCTGTTGCGGCGCTGGAAGAGGGTGGGTGATGCTGCTGTCACCCGCAATC 1500

Qy 1501 GCCAGCTGCTGCTGCGAACTCGCGACAGGACCAAGTCTGCTCGCGGGTGGACCGG 1560
 Db 1501 GCCAGCTGCTGCTGCGAACTCGCGACAGGACCAAGTCTGCTCGCGGGTGGACCGG 1560

Qy 1561 CGCGCGCGGGGCAACCGGCGCGTCCGGGCAACCGATGCGGTGGGAGTCCGCGAAGTCGC 1620
 Db 1561 CGCGCGCGGGGCAACCGGCGCGTCCGGGCAACCGATGCGGTGGGAGTCCGCGAAGTCGC 1620

Qy 1621 CGCGCATCCGCGGTAGACTTTGAGGACTGGTGAAGCGCGTCTGTCAGTCCCGTGGCAG 1680
 Db 1621 CGCGCATCCGCGGTAGACTTTGAGGACTGGTGAAGCGCGTCTGTCAGTCCCGTGGCAG 1680

Qy 1681 CGAGCGCTCGGCGCGGCGACGCGCCGATCCACTCCAGCTCCGCGACCTCGCCGACACGC 1740
 Db 1681 CGAGCGCTCGGCGCGGCGACGCGCCGATCCACTCCAGCTCCGCGACCTCGCCGCGC 1740

Qy 1741 GGAATTCAGCTGGGAAACAGSCCACTTCCACTCCAGCTCCGCGACCTCGCCGACACGC 1800
 Db 1741 GGAATTCAGCTGGGAAACAGSCCACTTCCACTCCAGCTCCGCGACCTCGCCGACACGC 1800

Qy 1801 CGAGTGGTTCAGGTGGCAGGTGTAGTTCAGCGTGGCCATCAGGAAGTGGTAGGCCACCG 1860
 Db 1801 CGAGTGGTTCAGGTGGCAGGTGTAGTTCAGCGTGGCCATCAGGAAGTGGTAGGCCACCG 1860

Qy 1861 GCACGCGTAGCCGCGTGGAGACGAGCATCGCGCTGACGCGAGGTCTCGGGGATCA 1920
 Db 1861 GCACGCGTAGCCGCGTGGAGACGAGCATCGCGCTGACGCGAGGTCTCGGGGATCA 1920

Qy 1921 GCACGCGTAGCTGTGCTGCTGCGTGGGAGAGCGGTGCGTGGTGGGCGAGGA 1980
 Db 1921 GCACGCGTAGCTGTGCTGCTGCGTGGGAGAGCGGTGCGTGGTGGGCGAGGA 1980

Qy 1981 CGGCGGTGACCGAGAACCGGTCACCGCTTCGCTATCCAGTTCACCGGGAACCGGTCCT 2040
 Db 1981 CGGCGGTGACCGAGAACCGGTCACCGCTTCGCTATCCAGTTCACCGGGAACCGGTCCT 2040


```
Db 4201 AAAAGCGCGCGCGGGTCCAGGGTCTCTCAGAGCTTAGACCGCCACCCGCCC 4260
Qy 4261 TGCCGCCAGCGCGGACAGCGCCGCCAGCTCCGCCGCTGTCGCTCCGCGCGCGAGGCG 4320
Db 4261 TGCCGCCAGCGCGGACAGCGCCGCCAGCTCCGCCGCTGTCGCTCCGCGCGCGAGGCG 4320
Qy 4321 AGCGACGGGTAGAGCGCGCGGTACC 4346
Db 4321 AGCGACGGGTAGAGCGCGCGGTACC 4346

RESULT 2
ADE14794
ID ADE14794 standard; DNA; 4346 BP.
XX AC
XX AC
XX ADE14794;
DT 29-JAN-2004 (first entry)
XX Streptomyces coelicolor partial genome sequence 2.
DE Streptomyces coelicolor partial genome sequence 2.
XX antibiotic-producing strain; antibiotic production; scbA gene; ScbR gene;
XX actinorhodin; undecylprodigiosin; gene; ds.
XX Streptomyces coelicolor.
XX Key Location/Qualifiers
FH complement(1199..2143)
FT /*tag= a
FT /product= "Streptomyces coelicolor SrbB protein"
FT /transl_except= (3149..3147, aa:Tyr)
FT 2261..2908
FT /*tag= b
FT /product= "Streptomyces coelicolor SrbR protein"
FT complement(3024..3812)
FT /*tag= c
FT /product= "Streptomyces coelicolor SrbA protein"
FT /transl_except= (3150..3145, aa:Met)
PN US2003124644-A1.
XX
XX 03-JUL-2003.
PD
XX
XX 23-OCT-2001; 2001US-00017471.
PF
XX
XX 23-OCT-2000; 2000US-0242561P.
PR
XX (TAKA/) TAKANO E.
PA (BIBB/) BIBB M J.
XX
XX Takano E, Bibb MJ;
PI
XX
XX WPI; 2003-810983/76.
DR P-PDB; ADE14791, ADE14792, ADE14793.
DR
XX
XX Modifying an antibiotic-producing strain of Streptomyces coelicolor or
PT Streptomyces lividans to increase or alter the timing of antibiotic
PT production in the strain, comprises functionally deleting in the strain
PT the scbA or ScbR gene.
XX
XX Claim 19; Fig 14; 33pp; English.
XX
XX This invention relates to the novel modification of an antibiotic-
CC producing strain of Streptomyces coelicolor or Streptomyces lividans to
CC increase or to alter the timing of antibiotic production in the strain.
CC The method comprises functionally deleting in the strain the scbA or ScbR
CC gene. The method is useful in increasing and altering the timing of
CC antibiotic production (especially actinorhodin and undecylprodigiosin) in
CC Streptomyces species, particularly Streptomyces coelicolor or
CC Streptomyces lividans. The present sequence is that of a region of the
CC Streptomyces coelicolor genome, which encodes the ScbA, ScbB and ScbR
CC proteins and which is related to the invention.
```

```
XX SQ Sequence 4346 BP; 650 A; 1561 C; 1519 G; 616 T; 0 U; 0 Other;
Query Match 100.0%; Score 4346; DB 10; Length 4346;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTGACGACGCGCTCGGGTTTCGACGCGCGACGCGGTACTCGTTCCCGGCCACCGGCGCACCG 60
Db 1 GTGACGACGCGCTCGGGTTTCGACGCGCGACGCGGTACTCGTTCCCGGCCACCGGCGCACCG 60
Qy 61 GGTCTGCGCTCGATGACCGCATCGAGACGCTCGGCTGGCGGTCTCTGATAGTGAGC 120
Db 61 GGTCTGCGCTCGATGACCGCATCGAGACGCTCGGCTGGCGGTCTCTGATAGTGAGC 120
Qy 121 GGGCCCGCGCGGCGACGACATCGAGTCCATCTCCACTGCTGCGGCCCGCGAAAGTGAGC 180
Db 121 GGGCCCGCGCGGCGACGACATCGAGTCCATCTCCACTGCTGCGGCCCGCGAAAGTGAGC 180
Qy 181 ACGCACCGCGGACGTGACGCCATGGGAGGGCCACGTCGCGGACGGATCACCCCTGGCT 240
Db 181 ACCGACCGCGGACGTGACGCCATGGGAGGGCCACGTCGCGGACGGATCACCCCTGGCT 240
Qy 241 TCGGCCGAAAGGCTTTCGCGTGGTCCGCCCGCCACAGATGCGGAAACGCGCTGGCGGCGCGCC 300
Db 241 TCGGCCGAAAGGCTTTCGCGTGGTCCGCCCGCCACAGATGCGGAAACGCGCTGGCGGCGCGCC 300
Qy 301 CGTCACTTCCGCGACGGTTCGCGACGACCTGCGGCTTGGCCCCCGCGCTCGCGCTCGGC 360
Db 301 CGTCACTTCCGCGACGGTTCGCGACGACCTGCGGCTTGGCCCCCGCGCTCGCGCTCGGC 360
Qy 361 GCTCTCAGGAAACGCTCGCGACGCGGCTCGGATCTTCGCGAGCATCTCTCGCGGCGC 420
Db 361 GCTCTCAGGAAACGCTCGCGACGCGGCTCGGATCTTCGCGAGCATCTCTCGCGGCGC 420
Qy 421 CGCTTCCAGCCCCAGCTCTCGAAACGACGATGACCGGCCACGACCTCGGAGAGCGCGGC 480
Db 421 CGCTTCCAGCCCCAGCTCTCGAAACGACGATGACCGGCCACGACCTCGGAGAGCGCGGC 480
Qy 481 CGTCTGCTGCTGGCGGTGATCGGCTCGGGCCCGGACGCGGTCACGCCCTGCTCTGTCG 540
Db 481 CGTCTGCTGCTGGCGGTGATCGGCTCGGGCCCGGACGCGGTCACGCCCTGCTCTGTCG 540
Qy 541 GCGGCGCGGGTCAGTGCCCTGACCGCCTTCGCGGATGTCGCGGATCGACGACGCGC 600
Db 541 GCGGCGCGGGTCAGTGCCCTGACCGCCTTCGCGGATGTCGCGGATCGACGACGCGC 600
Qy 601 AACCGGGACGTCGCGTACAGCGCGGACCAACGCGCTCGGACCGGATGGCGGGCGGCCA 660
Db 601 AACCGGGACGTCGCGTACAGCGCGGACCAACGCGCTCGGACCGGATGGCGGGCGGCCA 660
Qy 661 GGACAGCGTGTGGACATGAAGGTCCTGGCCCGCAGGAAGTCCAGTCTAGCCCGGACTC 720
Db 661 GGACAGCGTGTGGACATGAAGGTCCTGGCCCGCAGGAAGTCCAGTCTAGCCCGGACTC 720
Qy 721 GCGTACGCGCGCTGCTCTCGCGCTCGCGGCTGATGAAGTCTCGCGCGCGCGGTTTC 780
Db 721 GCGTACGCGCGCTGCTCTCGCGCTCGCGGCTGATGAAGTCTCGCGCGCGCGGTTTC 780
Qy 781 CCCACACCGCGAGCATGGAACAGCTTCAACAGGTGCGGACGCGCGGCTCTCGCGCGCGCGC 840
Db 781 CCCACACCGCGAGCATGGAACAGCTTCAACAGGTGCGGACGCGCGGCTCTCGCGCGCGCGC 840
Qy 841 CGCGAAACGCTGCTCGCTCGCGCTCGGTGCACTGTTCTGAGAGGAACGCGCGCGCGCAC 900
Db 841 CGCGAAACGCTGCTCGCTCGCGCTCGGTGCACTGTTCTGAGAGGAACGCGCGCGCGCAC 900
Qy 901 CCGGTTGAGGGCCCGGTCACGCGCGGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 901 CCGGTTGAGGGCCCGGTCACGCGCGGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Qy 961 GTTCCGGCGCGGTGACGCTGTCACCGCTTCCGGCGCGCGCGGAGGACTCTGAGCGGACCGGT 1020
Db 961 GTTCCGGCGCGGTGACGCTGTCACCGCTTCCGGCGCGCGCGGAGGACTCTGAGCGGACCGGT 1020
```

Db 961 GTTCGGGCGGTGACGGGTACCGTTCACCGTTCGGGCGCCGGGCGAGGACTCTGACGGGACCGGT 1020
Qy 1021 CCGGGCCAGCAGCTGGGGCGACCTGACGGCCGACCAACCGGTGACCGCGGTCAAGAAT 1080
Db 1021 CCGGGCCAGCAGCTGGGGCGACCTGACGGCCGACCAACCGGTGACCGCGGTCAAGAAT 1080
Qy 1081 CACTCGGGGCTCTCTCGGGCAGCAGGAGCAGGGGCGCTCCGAAATACATATGAGGGGA 1140
Db 1081 CACTCGGGGCTCTCTCGGGCAGCAGGAGCAGGGGCGCTCCGAAATACATATGAGGGGA 1140
Qy 1141 AGGGCAGGATCTGCCCCGGGGCGGAAACGGGCGATGTTGCGGCCCGGGGCGGTGCTTC 1200
Db 1141 AGGGCAGGATCTGCCCCGGGGCGGAAACGGGCGATGTTGCGGCCCGGGGCGGTGCTTC 1200
Qy 1201 AGCCGAGAAACCGGGGCGGACAGCGTGGTGAAGAAAGAGGGTGCCTCTGATGCC 1260
Db 1201 AGCCGAGAAACCGGGGCGGACAGCGTGGTGAAGAAAGAGGGTGCCTCTGATGCC 1260
Qy 1261 CGGTGACCCGACAGGTGTCAGCCCGCGCCCGCGGCCCGGGCCGACCGTGGCTGATGCC 1320
Db 1261 CGGTGACCCGACAGGTGTCAGCCCGCGCCCGCGGCCCGGGCCGACCGTGGCTGATGCC 1320
Qy 1321 AGCAGGGGTGCGAACTCCGGGTACCGGACCGAAACCGGGTGCCTGATCGACGACGA 1380
Db 1321 AGCAGGGGTGCGAACTCCGGGTACCGGACCGAAACCGGGTGCCTGATCGACGACGA 1380
Qy 1381 AGGGCGCGGACCGGTGACAGGCGCGCTGCGGTGCGCGCTCGAGCAGCATGC 1440
Db 1381 AGGGCGCGGACCGGTGACAGGCGCGCTGCGGTGCGCGCTCGAGCAGCATGC 1440
Qy 1441 CGGGTACGTGTCGTTGGGGCGCTGGAAGAGGGTGGGTGACTGGTGTCCACCCGAGTC 1500
Db 1441 CGGGTACGTGTCGTTGGGGCGCTGGAAGAGGGTGGGTGACTGGTGTCCACCCGAGTC 1500
Qy 1501 GCGAGTGTCTGTCGGAATTCGCGGACAGGACGACGTCCTGTCGCGGGTGGACCGG 1560
Db 1501 GCGAGTGTCTGTCGGAATTCGCGGACAGGACGACGTCCTGTCGCGGGTGGACCGG 1560
Qy 1561 CCGCGCGCGGGGACGGGCGGGTCCCGGGCACCGATCGGTGGAGTCGCGAGTGC 1620
Db 1561 CCGCGCGCGGGGACGGGCGGGTCCCGGGCACCGATCGGTGGAGTCGCGAGTGC 1620
Qy 1621 CCGCATCCCGCGGTAGACTTGAGGACTGTTGAGGACTGTTGAGGCGCTGTTGAGGCG 1680
Db 1621 CCGCATCCCGCGGTAGACTTGAGGACTGTTGAGGACTGTTGAGGCGCTGTTGAGGCG 1680
Qy 1681 CGAGCGTCCGGCGGGCGACGGGCCAGTCCACCTGTCCTGTCAGGGTGCCTGCGCG 1740
Db 1681 CGAGCGTCCGGCGGGCGACGGGCCAGTCCACCTGTCCTGTCAGGGTGCCTGCGCG 1740
Qy 1741 GGAACTTCAGCTGGGAAAGAGGCGCTTCCAGCTCCAGCTCCGCGACTCGCCGACACGC 1800
Db 1741 GGAACTTCAGCTGGGAAAGAGGCGCTTCCAGCTCCAGCTCCGCGACTCGCCGACACGC 1800
Qy 1801 CGAGGTGGTTCAGGTGCGAGTGTAGTCCAGCGTGGCCATCAGGAAAGTGGTAGCCACCG 1860
Db 1801 CGAGGTGGTTCAGGTGCGAGTGTAGTCCAGCGTGGCCATCAGGAAAGTGGTAGCCACCG 1860
Qy 1861 GCAAGCGGTAGCGGGGTGGAAGACAGAGCATCCCGCTGACCGAGGCTCTCGGCGATCA 1920
Db 1861 GCAAGCGGTAGCGGGGTGGAAGACAGAGCATCCCGCTGACCGAGGCTCTCGGCGATCA 1920
Qy 1921 GCAGCGGATCGTTCGGTCCCGGTGACCGGTGCGAAGAACCGGTGGTCTGGGGCAGGA 1980
Db 1921 GCAGCGGATCGTTCGGTCCCGGTGACCGGTGCGAAGAACCGGTGGTCTGGGGCAGGA 1980
Qy 1981 CCGCGGTGACCGGAAACCGGTGACCGGCTTCCGATCCAGCTGACCGGGAAACCGGTCTCT 2040
Db 1981 CCGCGGTGACCGGAAACCGGTGACCGGCTTCCGATCCAGCTGACCGGGAAACCGGTCTCT 2040
Qy 2041 GCACCTGTCGGTGGACAGCGCCATCGGAACCGGGCTATCGGATCGAGT 2100
Db 2041 GCACCTGTCGGTGGACAGCGCCATCGGAACCGGGCTATCGGATCGAGT 2100

Qy 2101 TGGCATCGGACGCAAGATTGATCAAAAATACTGCTTCGGGCGATGGGTCCCCCCAGGAAT 2160
Db 2101 TGGCATCGGACGCAAGATTGATCAAAAATACTGCTTCGGGCGATGGGTCCCCCCAGGAAT 2160
Qy 2161 CATGTGATCGCGAGCTGTTCTGTATGCGGAACTTTAGATACAGACTGAGCGGTTTTTT 2220
Db 2161 CATGTGATCGCGAGCTGTTCTGTATGCGGAACTTTAGATACAGACTGAGCGGTTTTTT 2220
Qy 2221 TTCTATCTTCCCGGGGAGACATGAACAAGAGGACGAGCATGGCCAAAGCAGGACCGGC 2280
Db 2221 TTCTATCTTCCCGGGGAGACATGAACAAGAGGACGAGCATGGCCAAAGCAGGACCGGC 2280
Qy 2281 GATTCGCAAGCGGCGAGACGATCTGGAACGCGCGGGCGAGGTCTTCGAGAGCAGGCGCTA 2340
Db 2281 GATTCGCAAGCGGCGAGACGATCTGGAACGCGCGGGCGAGGTCTTCGAGAGCAGGCGCTA 2340
Qy 2341 CCAAGCTCCAGATCACGAGATCTCAAGGTGCGCGGGTGACCAAGGAGAGCCCTCTA 2400
Db 2341 CCAAGCTCCAGATCACGAGATCTCAAGGTGCGCGGGTGACCAAGGAGAGCCCTCTA 2400
Qy 2401 CTTTCACTTCCAGTCCAAAGGAGAACTGGCGCTGGGCGTCTTCGACGCGCCAGGAACACC 2460
Db 2401 CTTTCACTTCCAGTCCAAAGGAGAACTGGCGCTGGGCGTCTTCGACGCGCCAGGAACACC 2460
Qy 2461 ACAGGCGGTTCCGAGGCAACCCCTCCGCTCGGCTGCAAGAACTCATCGACATGGGATGTT 2520
Db 2461 ACAGGCGGTTCCGAGGCAACCCCTCCGCTCGGCTGCAAGAACTCATCGACATGGGATGTT 2520
Qy 2521 CTGTCAACGCTTGGCGACGACGTCGCGCGCGCGCGCGCTCTCATGAGACCA 2580
Db 2521 CTGTCAACGCTTGGCGACGACGTCGCGCGCGCGCGCGCTCTCATGAGACCA 2580
Qy 2581 GCAGCGCAGCTGCTGATCGCGGAGACCTTCGCTCGCTGGCGACGAGACACTCTCGAA 2640
Db 2581 GCAGCGCAGCTGCTGATCGCGGAGACCTTCGCTCGCTGGCGACGAGACACTCTCGAA 2640
Qy 2641 GCTGCTGAACCAAGGCAACGGTGAAGTGTGCTGCCCATGTGCTCAACCGACTC 2700
Db 2641 GCTGCTGAACCAAGGCAACGGTGAAGTGTGCTGCCCATGTGCTCAACCGACTC 2700
Qy 2701 GCGCGATCTTACGTGGGCAAGTTCGCGGGATACAGGTCTGTCAGAGCGTCAAGCA 2760
Db 2701 GCGCGATCTTACGTGGGCAAGTTCGCGGGATACAGGTCTGTCAGAGCGTCAAGCA 2760
Qy 2761 CTACAGGACCTCGAACACCGCTACGCGCTGCTGAGAAAGCAGACATCTGCGCGCATCGC 2820
Db 2761 CTACAGGACCTCGAACACCGCTACGCGCTGCTGAGAAAGCAGACATCTGCGCGCATCGC 2820
Qy 2821 GGTTCCTCGTGGCGCGCTCGATCTCTCCGAGGAGCGCGAGCAGCGCTCGCGC 2880
Db 2821 GGTTCCTCGTGGCGCGCTCGATCTCTCCGAGGAGCGCGAGCAGCGCTCGCGC 2880
Qy 2881 CGAACTGGCACCGACCGGAGAGACTGACCGCGGAGCGCGCGCGCGCTAGAGTCTGC 2940
Db 2881 CGAACTGGCACCGACCGGAGAGACTGACCGCGGAGCGCGCGCGCGCTAGAGTCTGC 2940
Qy 2941 CCGTGGCCGAGCGCGACCGCGCGCTACGGGCGCGCGCGCGCGCGCTAGAGTCTGC 3000
Db 2941 CCGTGGCCGAGCGCGACCGCGCGCTACGGGCGCGCGCGCGCGCGCTAGAGTCTGC 3000
Qy 3001 CTTGCGTACCGAAGCGTGGCGGTGAGAAATGCTTCGCTGCGCTGCGATCGAGTCTGG 3060
Db 3001 CTTGCGTACCGAAGCGTGGCGGTGAGAAATGCTTCGCTGCGCTGCGATCGAGTCTGG 3060
Qy 3061 CCGGTGATCCACGCTGAGTTCGAGGCGGAGGCGGAGGCGGAGTCTGGCG 3120
Db 3061 CCGGTGATCCACGCTGAGTTCGAGGCGGAGGCGGAGGCGGAGTCTGGCG 3120
Qy 3121 GGTCTGCGGATCGGTTGAACACGAGTTCGGGCGGAGTTCGGCGCGGCTCGGGGCTC 3180
Db 3121 GGTCTGCGGATCGGTTGAACACGAGTTCGGGCGGAGTTCGGCGCGGCTCGGGGCTC 3180

QY 3181 TGGCGCGCGTGGTTCATGCTCGTCTCCACGAAACCGCGCGCCACCGCGCTTGACCGTG 3240
Db |||||
QY 3181 TGGCGCGCGTGGTTCATGCTCGTCTCCACGAAACCGCGCGCCACCGCGCTTGACCGTG 3240
Db |||||
QY 3241 ATCCCCCGTTCCCGCAGTTGCTTGCCACGAGCGCGAGCGTGAGCGTGTTCCACGCGCACCCCTTG 3300
Db |||||
QY 3241 ATCCCCCGTTCCCGCAGTTGCTTGCCACGAGCGCGAGCGTGAGCGTGTTCCACGCGCACCCCTTG 3300
Db |||||
QY 3301 GTCATCGGCTATCCGATGGACTCGGCGGAACCGCGCGCCCGGTCGCGGCGAGCAGATGTTG 3360
Db |||||
QY 3301 GTCATCGGCTATCCGATGGACTCGGCGGAACCGCGCGCCCGGTCGCGGCGAGCAGATGTTG 3360
Db |||||
QY 3361 ATGATCCGCGCGCGTGCAGCTGTTTTCAGTCCGTCGACCCAGGAAACAGCGGTGCC 3420
Db |||||
QY 3361 ATGATCCGCGCGCGTGCAGCTGTTTTCAGTCCGTCGACCCAGGAAACAGCGGTGCC 3420
Db |||||
QY 3421 CGGACGTTGACGGCGACCGAGTCGGTTCGAAGACCTCTCGGTGACTTCCGTTGATCCGTCGCC 3480
Db |||||
QY 3421 CGGACGTTGACGGCGACCGAGTCGGTTCGAAGACCTCTCGGTGACTTCCGTTGATCCGTCGCC 3480
Db |||||
QY 3481 GAGCGCTGACGCCCGCTTGTTCACAGGATGTCGAACCTCGGCGGCACTCCGAACTCG 3540
Db |||||
QY 3481 GAGCGCTGACGCCCGCTTGTTCACAGGATGTCGAACCTCGGCGGCACTCCGAACTCG 3540
Db |||||
QY 3541 CCCATCCGCGGTGGAACCGCGGTAGAGCGGCGGTCCACCCAGACCGCGAGTTTCG 3600
Db |||||
QY 3541 CCCATCCGCGGTGGAACCGCGGTAGAGCGGCGGTCCACCCAGACCGCGAGTTTCG 3600
Db |||||
QY 3601 GCCCGGATGGCAACCGCTGTCGCGCGTGTCTCGGATGCTCTCGACGCTCTCTCGCGCC 3660
Db |||||
QY 3601 GCCCGGATGGCAACCGCTGTCGCGCGTGTCTCGGATGCTCTCGACGCTCTCTCGCGCC 3660
Db |||||
QY 3661 GCCGCTCGCTGCTGCTAGTGAATGCAAGAGCGGCGGTCGCGGCGGCGAGCG 3720
Db |||||
QY 3661 GCCGCTCGCTGCTGCTAGTGAATGCAAGAGCGGCGGTCGCGGCGGCGAGCG 3720
Db |||||
QY 3721 GCGATACCGGTCGATCCCGCGTTCGCCCGTCCACGAGCGGCTCTTCCCTCCAGC 3780
Db |||||
QY 3721 GCGATACCGGTCGATCCCGCGTTCGCCCGTCCACGAGCGGCTCTTCCCTCCAGC 3780
Db |||||
QY 3781 GGTCTTCCATACCTCGTCCCATGTGTCAGCATATCAGCCCGCGCGTGTGAGCGACCC 3840
Db |||||
QY 3781 GGTCTTCCATACCTCGTCCCATGTGTCAGCATATCAGCCCGCGCGTGTGAGCGACCC 3840
Db |||||
QY 3841 ATGGCGCGCGTTCGCGCTTGCATTCGACGGTTCACAGCTACTGTGACCGCGTCAGACG 3900
Db |||||
QY 3841 ATGGCGCGCGTTCGCGCTTGCATTCGACGGTTCACAGCTACTGTGACCGCGTCAGACG 3900
Db |||||
QY 3901 GGGCGGAGTGGCGGTTGGACGGCTGGGGCGAGATCGGGCGGCGCGCACGGGAAACCG 3960
Db |||||
QY 3901 GGGCGGAGTGGCGGTTGGACGGCTGGGGCGAGATCGGGCGGCGCGCACGGGAAACCG 3960
Db |||||
QY 3961 GCGCGGTCAGGGGTCAGGGGTCGCGGGGACCGCCAGCGCGGTTCAGGGCACCGACCGGA 4020
Db |||||
QY 3961 GCGCGGTCAGGGGTCAGGGGTCGCGGGGACCGCCAGCGCGGTTCAGGGCACCGACCGGA 4020
Db |||||
QY 4021 TCGAGGTTCGGGTTCGACGCGGCAACGAGTCTTCGCGGCGGAGCTTCGATTCGTCGCG 4080
Db |||||
QY 4021 TCGAGGTTCGGGTTCGACGCGGCAACGAGTCTTCGCGGCGGAGCTTCGATTCGTCGCG 4080
Db |||||
QY 4081 TACAGAGCGGTCGCGCGGAGTCTGAGCTGAGCTGCGCGCGGCTGGGTGAGGCGG 4140
Db |||||
QY 4081 TACAGAGCGGTCGCGCGGAGTCTGAGCTGAGCTGCGCGCGGCTGGGTGAGGCGG 4140
Db |||||
QY 4141 TTGGCCACGAGGCGGAAAGCGGGAGGTTCGCGCGGAGCATCATGGGCGGCGCGGTTCG 4200
Db |||||
QY 4141 TTGGCCACGAGGCGGAAAGCGGGAGGTTCGCGCGGAGCATCATGGGCGGCGCGGTTCG 4200
Db |||||
QY 4201 AAAAGCGCGGCGGCGGTCGAGGGCTCTTCAGGACGTCTAGACCGCGCAACCGCGCC 4260
Db |||||
QY 4201 AAAAGCGCGGCGGCGGTCGAGGGCTCTTCAGGACGTCTAGACCGCGCAACCGCGCC 4260
Db |||||
QY 4261 TGGCGCAGGCGGCGACGCGCGCGCCAGCTCCGCGGTGTCTCGCTCCGCGCGCGGCGG 4320

Db 4261 TGGCGCGCGCGCGGCGAGCGGCCCGCGCGCTCCCGCTGTGCGCTCCGCGCGCGGAGCG 4320
QY 4321 AGCGACGCGTAGAGCGCGCGGTACC 4346
Db 4321 AGCGACGCGTAGAGCGCGCGGTACC 4346
RESULT 3
ABN86377
ID ABN86377 standard; DNA; 401 BP.
XX
AC ABN86377;
XX
DT 21-OCT-2002 (first entry)
XX
DE S. coelicolor ScbA and ScbR proteins encoding DNA fragment.
XX
KW Antibiotic; bacterium; scbA; afsA; scbR; arpA; barA; actinorhodine; Act;
KW undecylprodigiosin; Red; gene; ds.
XX
OS Streptomyces coelicolor.
XX
FH Key Location/Qualifiers
FT CDS complement(1..171)
FT /*tag= a
FT /product= "ScbA fragment"
FT /partial
FT /note= "the stop codon is not indicated"
FT 289..399
FT /*tag= b
FT /product= "ScbR fragment"
FT /partial
FT /note= "the stop codon is not indicated"
XX
FN CA2322241-A1.
XX
PD 23-APR-2002.
XX
PF 23-OCT-2000; 2000CA-02322241.
XX
PR 23-OCT-2000; 2000CA-02322241.
XX
PA (PLAN-) PLANT BIOSCIENCE LTD.
XX
PI Takano E, Bibb M;
XX
DR WPI; 2002-501089/54.
XX
P-PSDB; ABB80938, ABB80939.
XX
PT Modifying antibiotic-producing Streptomyces, to increase, or alter timing
PT of, antibiotic production, by deleting the scbA or scbR genes.
XX
PS Disclosure; Fig 4B; 64pp; English.
XX
CC The invention provides a method for modifying an antibiotic-producing
CC strain of Streptomyces to increase production of antibiotics or to alter
CC the timing of antibiotic production. The modification is functional
CC deletion of the scbA gene of S. coelicolor, or its homologues, but is not
CC deletion of the afsA gene of S. griseus, or the modification is
CC functional deletion of the scbR gene of S. coelicolor, or its homologues,
CC but is not deletion of arpA of S. griseus nor barA of S. virginiae. The
CC method is particularly used for production of the antibiotics
CC actinorhodine (Act) and undecylprodigiosin (Red). The present sequence
CC represents a DNA fragment encoding the S. coelicolor ScbA and ScbR
XX
SQ Sequence 401 BP; 87 A; 114 C; 128 G; 72 T; 0 U; 0 Other;
Query Match 9.2%; Score 401; DB 6; Length 401;
Best Local Similarity 100.0%; Pred. No. 1.3e-53;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1973 GGGCAGGACGGCGGTGACCGAGAACCGGTCAACCGCCCTTCGGTATCCAGCTGACCGGGAA 2032
DB 1 GGGCAGGACGGCGGTGACCGAGAACCGGTCAACCGCCCTTCGGTATCCAGCTGACCGGGAA 60
QY 2033 CGCGTCTCGACCCCTGGTCCGCTGACAGCGCCATCGGACCGCAATGCGGTTTGTTC 2092
DB 61 CGCGTCTCGACCCCTGGTCCGCTGACAGCGCCATCGGACCGCAATGCGGTTTGTTC 120
QY 2093 GATCGAGTTGGCATCGGACGAGAAATTGATCAAAACTACTGCTTCGGGCATGGGTCCCCC 2152
DB 121 GATCGAGTTGGCATCGGACGAGAAATTGATCAAAACTACTGCTTCGGGCATGGGTCCCCC 180
QY 2153 CCAGGAATCATGTGATGCGGAGCTGTTCTGTATGCGCGAACCTTAAGATACAGACTGAGC 2212
DB 181 CCAGGAATCATGTGATGCGGAGCTGTTCTGTATGCGCGAACCTTAAGATACAGACTGAGC 240
QY 2213 GGTGTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGAGGATGCCCAAGCAG 2272
DB 241 GGTGTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGAGGATGCCCAAGCAG 300
QY 2273 GACCGGGGATCCGCAACCGGCGAGACGATCCTTGGACGCCCGCGCGAGCTTTCGAGAAG 2332
DB 301 GACCGGGGATCCGCAACCGGCGAGACGATCCTTGGACGCCCGCGCGAGCTTTCGAGAAG 360
QY 2333 CAGGGCTACCAAGTCCACAGATCAGGAGATCCTCAAGGT 2373
DB 361 CAGGGCTACCAAGTCCACAGATCAGGAGATCCTCAAGGT 401

RESULT 4

ID ABN86378/c
ID ABN86378 standard; DNA; 401 BP.
AC ABN86378;
XX
XX
DT 21-OCT-2002 (first entry)
XX
XX
DE S. coelicolor ScbA and ScbR proteins encoding complementary DNA.
XX
XX
KW Antibiotic; bacterium; scbA; afeA; scbR; arpA; barA; actinorhodine; Act;
KW undecylprodigiosin; Red; gene; ds.
XX

Streptomyces coelicolor.

OS
XX
XX
PN CA2322241-A1.
XX
XX
PD 23-APR-2002.
XX
XX
PF 23-OCT-2000; 2000CA-02322241.
XX
XX
PR 23-OCT-2000; 2000CA-02322241.
XX
XX
PA (PLAN-) PLANT BIOSCIENCE LTD.
XX
XX
PI Takano E, Bibb M;
XX
XX
DR WPI; 2002-501089/54.
XX
XX
PT Modifying antibiotic-producing Streptomyces, to increase, or alter timing
PT of, antibiotic production, by deleting the scbA or scbR genes.
XX
XX
PS Disclosure; Fig 4B; 64pp; English.

XX
CC The invention provides a method for modifying an antibiotic-producing
CC strain of Streptomyces to increase production of antibiotics or to alter
CC the timing of antibiotic production. The modification is functional
CC deletion of the scbA gene of S. coelicolor, or its homologues, but is not
CC deletion of the afeA gene of S. griseus, or the modification is
CC functional deletion of the scbR gene of S. coelicolor, or its homologues,
CC but is not deletion of arpA of S. griseus nor barA of S. virginiae. The
CC method is particularly used for production of the antibiotics
CC actinorhodine (Act) and undecylprodigiosin (Red). The present sequence
CC represents a complementary fragment of the DNA encoding the S. coelicolor

CC ScbA and ScbR proteins
XX
SQ Sequence 401 BP; 72 A; 128 C; 114 G; 87 T; 0 U; 0 Other;
Query Match 9.2%; Score 401; DB 6; Length 401;
Best Local Similarity 100.0%; Pred. No. 1.3e-53;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1973 GGGCAGGACGGCGGTGACCGAGAACCGGTCAACCGCCCTTCGGTATCCAGCTGACCGGGAA 2032
DB 401 GGGCAGGACGGCGGTGACCGAGAACCGGTCAACCGCCCTTCGGTATCCAGCTGACCGGGAA 342
QY 2033 CGCGTCTCGACCCCTGGTCCGCTGACAGCGCCATCGGACCGCAATGCGGTTTGTTC 2092
DB 341 CGCGTCTCGACCCCTGGTCCGCTGACAGCGCCATCGGACCGCAATGCGGTTTGTTC 282
QY 2093 GATCGAGTTGGCATCGGACGAGAAATTGATCAAAACTACTGCTTCGGGCATGGGTCCCCC 2152
DB 281 GATCGAGTTGGCATCGGACGAGAAATTGATCAAAACTACTGCTTCGGGCATGGGTCCCCC 222
QY 2153 CCAGGAATCATGTGATGCGGAGCTGTTCTGTATGCGCGAACCTTAAGATACAGACTGAGC 2212
DB 221 CCAGGAATCATGTGATGCGGAGCTGTTCTGTATGCGCGAACCTTAAGATACAGACTGAGC 162
QY 2213 GGTGTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGAGGATGCCCAAGCAG 2272
DB 161 GGTGTTTTTCTATCTTCCCGGGGAGACATGAACAAGGAGGCGAGGATGCCCAAGCAG 102
QY 2273 GACCGGGGATCCGCAACCGGCGAGACGATCCTTGGACGCCCGCGCGAGCTTTCGAGAAG 2332
DB 101 GACCGGGGATCCGCAACCGGCGAGACGATCCTTGGACGCCCGCGCGAGCTTTCGAGAAG 42
QY 2333 CAGGGCTACCAAGTCCACAGATCAGGAGATCCTCAAGGT 2373
DB 41 CAGGGCTACCAAGTCCACAGATCAGGAGATCCTCAAGGT 1
RESULT 5
ADE14790
ID ADE14790 standard; DNA; 401 BP.
XX
AC ADE14790;
XX
XX
DT 29-JAN-2004 (first entry)
XX
XX
DE Streptomyces coelicolor partial genome sequence 1.
XX
XX
KW Streptomyces coelicolor partial genome sequence 1.
KW antibiotic-producing strain; antibiotic production; scbA gene; ScbR gene;
KW actinorhodin; undecylprodigiosin; gene; ds.
XX
OS Streptomyces coelicolor.
XX
XX
FH Key Location/Qualifiers
FT CDS complement(1..171)
FT /tag= a
FT /product= "Streptomyces coelicolor ScbA protein"
FT /partial
FT /note= "No stop codon"
FT CDS 289..399
FT /tag= b
FT /product= "Streptomyces coelicolor ScbR protein"
FT /partial
FT /note= "No stop codon"
XX
XX
PN US2003124644-A1.
XX
XX
PD 03-JUL-2003.
XX
XX
PF 23-OCT-2001; 2001US-00017471.
XX
XX
PR 23-OCT-2000; 2000US-0242561P.
XX
XX
PA (TAKA/) TAKANO E.

PA (BIBB/) BIBB M J.
XX
PI Takano E, Bibb MJ;
XX
DR WPI; 2003-810983/76.
XX P-PSDB; ADE14791, ADE14792.
XX
PT Modifying an antibiotic-producing strain of Streptomyces coelicolor or
PT Streptomyces lividans to increase or alter the timing of antibiotic
PT production in the strain, comprises functionally deleting in the strain
PT the scbA or ScbR gene.
XX
PS Example 4; Fig 4A; 33pp; English.
XX
CC This invention relates to the novel modification of an antibiotic-
CC producing strain of Streptomyces coelicolor or Streptomyces lividans to
CC increase or to alter the timing of antibiotic production in the strain.
CC The method comprises functionally deleting in the strain the scbA or ScbR
CC gene. The method is useful in increasing and altering the timing of
CC antibiotic production (especially actinorhodin and undecylprodigiosin) in
CC Streptomyces species, particularly Streptomyces coelicolor or
CC Streptomyces lividans. The present sequence is that of a region of the
CC Streptomyces coelicolor genome, which partially encodes the ScbA and ScbR
CC proteins and which is related to the invention.
XX
SQ Sequence 401 BP; 87 A; 114 C; 128 G; 72 T; 0 U; 0 Other;

Query Match 9.2%; Score 401; DB 10; Length 401;
Best Local Similarity 100.0%; Pred. No. 1.3e-53;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1973 GGGCAGGACGGGTGACCGAGAACCGGTACCGGCCCTTCGGTATCCAGCTGACCGGAA 2032
DB 1 GGGCAGGACGGGTGACCGAGAACCGGTACCGGCCCTTCGGTATCCAGCTGACCGGAA 60
QY 2033 CGCGTCTGCACTCCCTGCTCGGTGACCAAGCGCCATCGAACCGGCAATGCGTTTGTTC 2092
DB 61 CGCGTCTGCACTCCCTGCTCGGTGACCAAGCGCCATCGAACCGGCAATGCGTTTGTTC 120
QY 2093 GATCGAGTTGGCATCGGACGAGCAATTCATCAAACTACTGTTGGGCATGGGTCCCCC 2152
DB 121 GATCGAGTTGGCATCGGACGAGCAATTCATCAAACTACTGTTGGGCATGGGTCCCCC 180
QY 2153 CCAGGAATCATGTGATCGGAGCTGTTCTGTATGCGCGAAAGTTAAGATACAGACTGAGC 2212
DB 181 CCAGGAATCATGTGATCGGAGCTGTTCTGTATGCGCGAAAGTTAAGATACAGACTGAGC 240
QY 2213 GGTTTTTTTTCTATCTTTCCGGGGGAGACATGAACAAGGAGGCGATGCCCAAGCAG 2272
DB 241 GGTTTTTTTTCTATCTTTCCGGGGGAGACATGAACAAGGAGGCGATGCCCAAGCAG 300
QY 2273 GACCGGGCGATCCGCACGCGGAGACGATCTTGGACGCGCGGCGCAGGTCTTCGAGAAG 2332
DB 301 GACCGGGCGATCCGCACGCGGAGACGATCTTGGACGCGCGGCGCAGGTCTTCGAGAAG 360
QY 2333 CAGGGCTACCAAGCTGCCACGATCAGGAGATCTTCAAGGT 2373
DB 361 CAGGGCTACCAAGCTGCCACGATCAGGAGATCTTCAAGGT 401

RESULT 6
AAD47222
ID AAD47222 standard; DNA; 19016 BP.
XX
AC AAD47222;
XX
DT 24-FEB-2003 (first entry)
XX
DE Streptococcus sp. H021 DNA fragment for rabelomycin biosynthetic pathway.
XX
KW Gene cluster; rabelomycin biosynthetic pathway; angucycline compound;
KW drug screening; enzyme; gene; ds.
XX

OS Streptococcus sp.
XX
FH Key Location/Qualifiers
FT CDS complement(1..561)
FT /*tag= m
FT /product= "Orfs"
FT /note= "No stop codon"
FT /partial
FT complement(580..1335)
FT /*tag= n
FT /product= "OrfR"
FT complement(1391..2701)
FT /*tag= 1
FT /product= "OrfQ"
FT /transl_except= (pos:2699..2701, aa:Met)
FT complement(2712..3710)
FT /*tag= k
FT /product= "OrfH"
FT complement(3793..4848)
FT /*tag= s
FT /product= "Orf2"
FT complement(4854..5438)
FT /*tag= j
FT /product= "OrfO"
FT complement(5494..6687)
FT /*tag= o
FT /product= "OrfY"
FT complement(6780..8051)
FT /*tag= q
FT /product= "OrfJ"
FT complement(8178..8939)
FT /*tag= i
FT /product= "OrfV"
FT /transl_except= (pos:8937..8939, aa:Met)
FT complement(9014..10555)
FT /*tag= h
FT /product= "OrfM"
FT /transl_except= (pos:10553..10555, aa:Met)
FT complement(10557..11504)
FT /*tag= g
FT /product= "OrfL"
FT complement(11570..12352)
FT /*tag= d
FT /product= "OrfD"
FT /transl_except= (pos:12350..12352, aa:Met)
FT complement(12421..12684)
FT /*tag= c
FT /product= "OrfC"
FT complement(12681..13910)
FT /*tag= b
FT /product= "OrfB"
FT complement(13907..15142)
FT /*tag= a
FT /product= "OrfA"
FT /transl_except= (pos:15140..15142, aa:Met)
FT complement(15208..15537)
FT /*tag= f
FT /product= "OrfF"
FT complement(15586..17055)
FT /*tag= e
FT /product= "OrfE"
FT 17692..18492
FT /*tag= r
FT /product= "OrfI"
FT complement(18603..19016)
FT /*tag= p
FT /product= "OrfR1"
FT /note= "No start codon"
FT /partial
XX
PN WO200274800-A1.
XX
XX 26-SEP-2002.
PD

FT																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

QY	3257	GTTCGCTGGCCAGGGCGAGCGTGAGCGTGTCCACGCACCTTGGTCATCGCGTATCCGA	3316
DB	7345	CGTCTACGGCGAGGGCGGGGTGAGCACCTCGATCGCCCCCTTGGTCATGGCGTAGTGA	7286
QY	3317	TGGACTCGGGGAACCGCGCGGGTCGCGGCAGACGAGATTTGATGATCCGCCCGCGT	3376
DB	7285	TGACGTGCGGGTCTGCGCGTAGCGGCGGATCCGGAGGAAATGTTGACGATGACCCGCGT	7226
QY	3377	CGCGAGTCTGTTTCAGTCCGCTGCGGACAGGAAACAGGGGTGCCGAGACGTTGACGGGA	3436
DB	7225	CGTGCATCCGGGGCAGGGGCATGCCGAGTCAACAAAGAACGGGGCGCGCTGGTGTGAGTGGA	7166
QY	3437	CCAGTTCGTCGAAGACCTTCCTCGGTGACTTCCGTCCGAGCCGCGCTGACGCCCG	3496
DB	7165	CCAGACGTTCGAAATCTTCGACAGTGAAGCGCTCGAGCGCTCCGGGCGCATGATTCGG	7106
QY	3497	CGTTGTTTACCACGAGATGTGAACTCGGGCGGCACCTCCGAACTCCGCCATCCCGGCGTGA	3556
DB	7105	CGTTGTTGACGAGGATGTGAGGCCCTCGGTGAGCCCGGAGACCGTTCGAGCGCGGAGTGA	7046
QY	3557	ACGCGCGGTAGAGCGCGCGCGGTACCCACGACGCCGAGTTCGGCCCGCGATGGCCAAACG	3616
DB	7045	ACTCCCGCATCAGCTCTTTAGGCCCTCGGGTCGGGACAGGTTCGGCTTGGACCGGCGCG	6986
QY	3617	CCTGTCGCGCGCTGCTCGGATGTCGACGCTCTCTCGCGCGCGCGCTCGCTGCTGC	3676
DB	6985	CGAGGCCCGCGGACCGGTGATGCCGTGAC-ACGCGAGCGCGCGCGCTCCGATCGGATC	6927
QY	3677	CGTAGTGGACTGCCACGAGCGCCCCGTCGCGGCGCAGCCGCGGCGGATACCGGTCGGA	3736
DB	6926	CGTAGTGGACGATTACGCGGCGCACCGGAGCGCGCCACCGCATCTCGATGCGCGCCCGA	6867
QY	3737	TGCCCCCGGCTTCCCCCGGTACCCAGGGCG	3765
DB	6866	TGCGCGCGCTGGCCCCCGGTGACACGCGG	6838
RESULT 10			
ID	AEA81294 standard; DNA; 4721 BP.		
AC	AEA81294;		
DT	11-AUG-2005 (first entry)		
DE	Streptomyces natalensis natamycin regulator DNA SEQ ID NO 1.		
KW	biosynthetic regulator; natamycin; gene expression; gene regulation;		
OS	Streptomyces natalensis.		
PN	KR2005000441-A.		
PD	05-JAN-2005.		
PF	24-JUN-2003; 2003KR-00040918.		
PR	24-JUN-2003; 2003KR-00040918.		
PA	(BIOR-) BIO R & DS.		
PI	Bae DJ, Hwang YI, Jang HS, Kang SJ, Kim BJ, Lee CK, Lee KM;		
DR	WPI; 2005-352244/36.		
PT	Biosynthetic regulator of natamycin derived from streptomyces natalensis		
PT	used for mass-producing natamycin and nucleotide sequences encoding the		
PS	Example 2; SEQ ID NO 1; 21pp; Korean.		
CC	This invention describes a novel biosynthetic regulator of natamycin		

CC derived from Streptomyces natalensis (ATCC 27448) which can be used for
 CC regulating the biosynthesis of natamycin. The novel polypeptides and
 CC their encoding polynucleotides can be used for mass-producing natamycin
 CC in microorganisms using an expression vector containing the natamycin
 CC gene and culturing the transformed microorganism in a medium. This
 CC sequence represents a DNA fragment containing a biosynthetic regulator of
 CC natamycin.
 XX
 SQ Sequence 4721 BP; 690 A; 1464 C; 1704 G; 797 T; 0 U; 66 Other;
 Query Match 5.5%; Score 239; DB 14; Length 4721;
 Best Local Similarity 50.3%; Pred. No. 2,2e-28;
 Matches 876; Conservative 0; Mismatches 775; Indels 92; Gaps 8;
 QY 1213 CGGGGCGGACAGCGTGTGAGAAAGACGAGGCTGCGCTCTGATGCCCGGTGACCCGCA 1272
 DB 2678 CGGGGCGGTGAGCGTGTGAGAAACACCTTTGCGCGGTCTGTTGGCGGTGACGCGCA 2737
 QY 1273 CGGTGTGACG 1332
 DB 2738 CTGTGTGTGTGCGCGCTCACTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2794
 QY 1333 CGAACTCCCGGTACCGGACGAAACCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGAC 1392
 DB 2795 CGAACTTCGACGCTGCTGAAACACACCTTTGCGCGGTGAGCGCTGAAAGGTCTTTTGTG 2851
 QY 1393 CGGTACGAGGCGACG 1452
 DB 2852 CGGTACGAGGCG 2911
 QY 1453 CGTTGGGCGGTGAAAGAGGCTGCGGTGACGTGTGTGTCACCGCGCGCGCGCGCGCGCG 1512
 DB 2912 CGACGGGCGGTGCAACAGCGCGGTGTTGCGGAAGTTCGCGCGCGCGCGCGCGCGCG 2971
 QY 1513 GCTGCGAACTCGCGACAGACGACGACGACGACGACGACGACGACGACGACGACGACG 1572
 DB 2972 CGGTGTGCGGTGCGGCG 3031
 QY 1573 GCAAGCGCGCGGTGCGGCG 1629
 DB 3032 GAGCGGGGCG 3091
 QY 1630 GCGGTGTGACTTGAAGACTGTGAAAGCGCGGTGCGGTGAGTGTGCGGTGAGCGCGCGCG 1689
 DB 3092 GCGGTGTGACCCCTGCGGGAATCCATTTGAAACCGGCTGCGCGCGCGCGCGCGCGCG 3151
 QY 1690 CG 1749
 DB 3152 CGGCAAGTGTGATCTCATCTGTCGCGAGACCGGTGATGGAACCGCGCGCGCGCGCGCG 3211
 QY 1750 GCTGGGAACAGGCGCACTTCCACCTCCAGCTCCGCGACCTCGCGCGCGCGCGCGCGCG 1809
 DB 3212 TGATTTGCGAGACAAAGTCAAGAACCAAGGTGTGCGGTGCTTTTTCGCGCTCCAGCGCT 3271
 QY 1810 CGAGGTGGCAGGTGTGATCAGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1869
 DB 3272 CCAGTTCGAAGGCGAAGTTCGAAGTTCGAAGTTCGAAGTTCGAAGTTCGAAGTTCGAAG 3331
 QY 1870 AGCGCGGTGGAAGACGAGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1929
 DB 3332 AACTCGGTGCGCGAGCAGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3391
 QY 1930 CGGTGCGGTGCGCGAGCGGTGCGGAAGAACGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1989
 DB 3392 CAGCTGGAACCG 3448
 QY 1990 CGGAGAACCGGTGCG 2049
 DB 3449 CGCGGAAATTCATGTGCG 3508
 QY 2050 TCCGTGTGGAACAGGCGCATCGGAACCGGCAATCGGTTTGTTCGATCGAGTTGCGATCGG 2109
 DB 3509 GGAGGTGGACGAATTCCTTGGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3568

QY 2110 ACCGAGAAATTGATCAAAACTACTGCTTGGGCGATGGGTCCCCCGCAGGAATCATGTGATG 2169
 DB 3569 GGAGCGCGCAGGTTCCTGTTCGGTGGCAGAGGCTGTGCATACGCGCGAGCCCGCAGGATG 3628
 QY 2170 CCGAGCTGTTCTGTATGC-----GGGAACGTGA 2197
 DB 3629 GTGTGACATAGTTCCTCCCTCGGTACCTCAAGCAACAGAGCTTCGGATGGCTCCCAATA 3688
 QY 2198 AGATACAGACTGAGCGGTTTTTT-----TCT 2224
 DB 3689 AGATACGTAATATCGGTTTGTATTTTCAATGCGGTGCGCAGCGCGCTTCGGAACACTCG 3748
 QY 2225 ATCTTTCGCGGGGAGACATGAAACAGGAGGCA-----GGCATGCCAAGCAG 2272
 DB 3749 TTGCAACGAGGGGGGACGCGCATGCGCGGTAAACATGAGAGATTGTCTCATGGCAACAG 3808
 QY 2273 GACCGGGCGATCCGCAACGCGGAGCAGATCTCTGAGCGCGCGCGCGCGGTCTTCGAGAAG 2332
 DB 3809 GCACGGCGATACAGACGCGGCGGCGAGATCTCTGATGCGCGGGCTTGTGTTGACACAG 3868
 QY 2333 CAGGGCTACAAAGTGCACGATCAGCGAGATCTCTCAAGGTGCGCGGGGTGACCAAGGGA 2392
 DB 3869 CTCGGATACGAGTCCGCGACCATCACCGAGATCATCGCGAGCGCGCGGTGACCAAGGGG 3928
 QY 2393 GCGCTTACTTCCACTTCCAGTCCAGGAGAACTGGCGCTGGCGCTCTTCGACGCCAG 2452
 DB 3929 GCGCTGTACTTCCACTTCCAGTCCAGGAGAACTGGCGCGCGGTGCTCTTCGAGGAGGCC 3988
 QY 2453 GAACACACACAGCGCGTTCGCGAGCAACCCCTCGGCTGCAAGAACTCATCGACATGGGC 2512
 DB 3989 GTACACACTGAGGGGTGCTACCGCAGGACTCCAAACTGCGAGGAGTGGGTGTTGGG 4048
 QY 2513 ATGTTGTTTGTACCGCTTTCGCGACGAAAGTGTGCGCGCGCGCGCGCGGTGCGCTTCC 2572
 DB 4049 ATGGTGTTCGCGCACCGGTGCTCCCGCGAGCGATGCTCAGTCCCGCGCATCCGCTTCC 4108
 QY 2573 ATGACACGAGCGCGACCGGTCTGATCCGCGAGGACCTTCGCTGCGTGGCAGGAGACA 2632
 DB 4109 ACCGACTTGAAGCGCGGTGCTCTGTTTCGCGACCT---CCTGGACGCCCTGATCGCCTTC 4165
 QY 2633 CTCCTGAAGCTGTGAACGAGCGCAAGGAGAACGCTGAGTTGCTGCCCATGTGGTCACC 2692
 DB 4166 ATGGCAGACAGTGGCCCTTGGCCAAAGGAGCAGGCGGAACTGTGCGCACACCGTTCCG 4225
 QY 2693 ACCGACTCGCGCGATCTCTACGTTGGGACGTTTCGCGGAGATACAGTGTGTTCCAGACG 2752
 DB 4226 GAGGAACCGCGTGTGTCATCGTTCGCGGATGAGACCGGATGAGATCGTCACTCCGAC 4285
 QY 2753 GTGAGCGATACGAGGACCTCGAAGACCGGTACGCGCTGCTGCGAGGAGCATCTCTGCC 2812
 DB 4286 CTCAGGAGCGGCGAGTCCCTCCAGCACCGGGTTTTCAGGCTCTACGACCATCTTCTCCCC 4345
 QY 2813 GGCATCGCGTTCCCTTCGCTGCGCGCGCTCGATCTCTCCGAGGAGCGCGGAGCACGC 2872
 DB 4346 AGCATCGCGTTCGCGGTGCTGCGCGAGCTGAGTGCACCTCACCGGACCGCGGTACCCGG 4405
 QY 2873 CTC 2875
 DB 4406 ATC 4408
 RESULT 11
 ID ADW38421/C
 XX ADW38421 standard; DNA; 936 BP.
 AC ADW38421;
 XX
 DT 24-MAR-2005 (first entry)
 XX
 DE DNA sequence related to Streptomyces nanchangensis.
 XX
 KW Streptomyces nanchangensis; ds.

```
XX Streptomyces nanchangensis.
OS
XX Key Location/Qualifiers
FH 1..936
FT CDS
FT /*tag= a
XX
XX CN1472322-A.
FN
XX
XX 04-FEB-2004.
PD
XX
XX 16-MAY-2003; 2003CN-00116989.
PP
XX
XX 16-MAY-2003; 2003CN-00116989.
PR
XX
XX (UYSH-) UNIV SHANGHAI COMMUNICATION.
PA
XX Deng Z, Sun Y, Zhou X;
PI
XX WPI; 2004-317409/30.
DR
XX P-PSDB; ADM38422.
DR
XX
XX Meilin mould C5 keto reductases gene, useful for reducing the C5-keto
XX group of the Meilingmycin glycoside.
XX
XX Claim 1; Page 2; 9pp; Chinese.
XX
XX The present sequence is a DNA sequence related to Streptomyces
XX nanchangensis.
XX
XX Sequence 936 BP; 153 A; 328 C; 308 G; 147 T; 0 U; 0 Other;
SQ
Query Match 5.5%; Score 238; DB 13; Length 936;
Best Local Similarity 59.7%; Pred. No. 3.4e-28;
Matches 472; Conservative 0; Mismatches 285; Indels 33; Gaps 3;
QY 3014 CGGTGGCGGTGACAGATCGTTCCGCTGTGGCATCGACGTACTGCGCGGTGATCCACC 3073
DB 925 GTGGTGTGGGGCCAGGCGCATGCCCGCGAGGCGTCGACAGATGTGCCGTGACCCAGC 866
QY 3074 GTGAGTCTGCGAGGCCAGAAAGGCCACCATGCTCGCGGATGTCTCGGGTCTGCCGATGC 3133
DB 865 GGGCTCTGCGAGCGAGAGGCGGACGACATCGCGGATGTGCGACGGCTGACCGAGTC 806
QY 3134 GGTGAAACAGAGATTGGCGGCAATGTCGCGCGCGCTGCGGGGTCTGCGCGCGCGCGTG 3193
DB 805 GCGCAGTGCAGTATGGCGGCGACGCGCTGCGCGCGCGCGCATATCGTGACGCCACG 746
QY 3194 CGTTTCATGTCG ---TCTCCAGMAACCGCGCGCCACCGCTGACCGTGATCCCGCGTT 3250
DB 745 TGTGGTGTGGTGTGTCGTTGATGCGCGCGCGCATGTTGTGACGGTGATCCCGCGT 686
QY 3251 CCCCCAGTTGCTGCGCCAGGCGAGCGTGAGCGGTGTCCACCGCACCTTTGGTTCATCGCGT 3310
DB 685 GTCCAGCTCATTTGGCCAGCGCTGGCGCATGATCTCGAGGGCGCCTTGTCTATGGCGT 626
QY 3311 ATCCGATGATCTCGGGGAAACCGCGCGCGGTGCGCGGAGACGAGATGTTGATGATCCGCC 3370
DB 625 A---GGGCGACGACGCGCGCGAGCGATCCTGGTGACGCGGAGAGATGTTGACGATCCGTC 569
QY 3371 CGCGTTCGCGAGTGTGTTTCAGTTCGTGTCGACAGCAGACGCGGTCCCGAGCTTGA 3430
DB 568 CGCGTTCGCGAGGAGCGCGGACACACCGTTGGTGACGAAACCGGGGACGTCATTTGA 509
QY 3431 CGGCGACAGTCCGTGCGAAGACCTCTCGGTGACTTCCGTGATTCGCGTCCCGAGCGCGTGA 3490
DB 508 CGCTGAAGTACGGTGCAGACTCTCTGATGACCTCGGAATGCGCGCATAGCCGT 449
QY 3491 CGCCCGCGTGTGTTACAGGATGTGAACTCGGGCGGCGACTCCCGAAT----- 3538
DB 448 CGCGTTCGCGCGCGCGGTGCTCGCGGTGCTAGATGCGCGGTGTTGTTACACGGA 389
QY 3539 -----CGCCCATCCCGCGGTGCGAAGCGCGGTGAGAGCGCGCGCGGTGCAC 3583
```

```
DB 388 TGTTCAGCGCGCGCGGAGCCCGTCCGAGCTCGTGAAGCGGTGTCGAGGGCGT 329
QY 3584 CCACGACGCGAGTTTCGCGCCCGGATGCGCAACGCTGTTCGCCCGCTGTCGCGATGTTCT 3643
DB 328 CGTCCGCGCAAGGTTCGCGCCCTGATGCGAATGCTGCGCCCGCGCTGCTCGATCGACG 269
QY 3644 CGACGTTCTTCGCGCGCGCGCGCTCGTCTGCTGCGGTAGTGGACTGCCACGAGCGCCCGT 3703
DB 268 CGACCGTTTCTTCGCGGTTCGCGCTTTTGACCGCCGTAGTCACGCGCACACGACACCTT 209
QY 3704 CCGCGCGCAGCGCGGCGGATACCGGCTCGATGCGCCCGCGCTTCCCGCGTCCACGAGG 3763
DB 208 CGGCGCGAGCCTTCGCGGCGATGCGCGCGCGATGCGCGCGCGCGCGCGCGTCCACGAGT 149
QY 3764 CGGTCTTGCC 3773
DB 148 CGGTTTTC 139
RESULT 12
AAL44719
ID AAL44719 standard; DNA; 699 BP.
XX
AC AAL44719;
XX
XX 03-MAY-2002 (first entry)
DT
XX
DE S virginiae butanolide binding protein coding sequence.
XX
KW Butanolide binding protein; gene expression induction; operator;
XX transgenic plant; antibiotic production; gene; ds.
OS Streptomyces virginiae.
FH Key Location/Qualifiers
FT CDS 1..699
FT /*tag= a
FT /product= "butanolide binding protein"
XX
FN WO200196581-A1.
XX
XX 20-DEC-2001.
PD
XX
PF 15-JUN-2001; 2001WO-JP005096.
XX
PR 15-JUN-2000; 2000JP-00180466.
XX
XX (KANF ) KANEKA CORP.
PA
XX Shinmyo A, Kato K, Yamada Y, Nihira T, Shindo T;
PI
XX WPI; 2002-098073/13.
DR
XX P-PSDB; AAM48990.
XX
PT Inducing expression of gene under regulation by operator at actinomycetes
PT self-regulator provision site, useful in producing transformant tobacco
PT for production of antibiotics e.g. virginiamycin.
XX
XX Example 1; Page 46-48; 57pp; Japanese.
PS
XX
CC The present invention relates to a method of inducing the expression of a
CC gene under the regulation of an operator at an actinomycetes self-
CC regulator provision site comprising imparting the characters of a
CC repressor and an operator constituting a gene expression-inducing system
CC with use of the actinomycetes self-regulator as inducer to a plant by
CC gene transfer thereby providing the self-regulator to the transformed
CC plant..The method is useful in producing transformant tobacco for
CC production of antibiotics e.g. virginiamycin. The present sequence is the
CC Streptomyces virginiae butanolide binding protein coding sequence which
CC was used in the exemplification of the invention
XX
XX Sequence 699 BP; 116 A; 233 C; 253 G; 97 T; 0 U; 0 Other;
```

Query Match	5.3%	Score 229;	DB 6;	Length 699;
Best Local Similarity	60.8%;	Pred. No. 8.7e-240;	Indels 0;	Gaps 0;
Matches 373;	Conservative	0;	Mismatches 240;	
QY	2270	CAGGACCGGCGCATCCGACCGCGCAGACCATCTGTGACGCCGCGCGCAGGCTCTTCGAG	2329	
DB	34	CAGGAACGGGCGCTCCGACACCGCGCAGGCGATCGTGGGGCAGCGGCTCTGGTCTTCGAC	93	
QY	2330	AAGCAGGGCTACCAAGCTGCCACGATCACCGAGATCTCTAAGTGGCGGGGTGACCAAG	2389	
DB	94	GAGTACGGGTTCAGAGCGCGCACAGTGGCAGAGATCTCTCGGGGGCTCGGTACACCAAG	153	
QY	2390	GGAGCCCTCTACTTTCACATTCCAGTCCAAAGGAAGAACTGGCGCTGGGCGCTCTTCGACGCC	2449	
DB	154	GGCGCGATGTACTTTCACATTTCGCTTCCAAGGAGAGCTGGCCCGCGCGCTGGCCCGAG	213	
QY	2450	CAGGAACCAACCAAGGCGGTTCCGGAGCAACCCCTCCGGCTGCMAGAACTCATATGCAATG	2509	
DB	214	CAGACCTGACAGTGGCGGTGCGGGAATCCGGCTCCAAAGGCGCAGGAACTGGTAGACCTC	273	
QY	2510	GGCATGTGTGTTCTCATACCGCTTTCGGCACAGAACGTCTGTGGCCCGGGCCGGCTGCGGCTC	2569	
DB	274	ACCATGTGTGTGCGCCACGGGATGCTGCACATCCGATCTCTGCGGGCGGGACACGGGCTC	333	
QY	2570	TCCATGACACAGCAGCGCACGGTCTCGATCGCCGAGGACCTTTCGTCGCTGSCACAG	2629	
DB	334	GCATCGACAGGGGCGGTGGATTTCTTCGACGCCAACCGTTTCGCGGAGTGGGGCGAC	393	
QY	2630	ACACTCTGAAGCTGTGAACCAAGCCAGGACAGGAGAACGGTGAAGTTGTCTGCCCATGTGCTC	2689	
DB	394	ATCTGCGCCAGCTCTCTGGCGGAGGCACAGGAACGGGGGAGGTGCTTCGCGACGTGAAC	453	
QY	2690	ACCAACGACTTCGGCCGATCTCTAGCTGGGACAGTTTCGCCGGGATACAGGTGCTGTCCCG	2749	
DB	454	CCGAAAGAACCGCGGACTTCATGTCGGTGTCTTACCGGGCTCCAGGGGCTCTCCCGG	513	
QY	2750	ACGGTGAAGCACTACAGGACCTCGAACAACCGCTACCGGCTGTCTGACAGAGCACTCTTG	2809	
DB	514	GTCACTTCGACCGCCAGGACCTCGGCCACCGGATCTCGGTGATGTGGAAACCAACGCTGCTG	573	
QY	2810	CCGCCATCGGTTTCCCTCGTCTGGCGCGCTCGATCTCTCCGAGGAGCGCGGAGCA	2869	
DB	574	CCAGCATGTGCCGGGTTCATGCTGACTGGATCGAAACCGCGGAGAGCGGATCGGG	633	
QY	2870	CGCTTCGCGCCG	2882	
DB	634	AAGTCTCGCGCGG	646	

RESULT 13
ACL64643/c
ID ACL64643 standard; DNA; 14638 BP.
XX
AC ACL64643;
XX
XX 02-JUN-2005 (first entry)
XX
XX DT
XX
DE M. xanthus DNA fragment, seq id 1106.
XX
XX Transgenic plant; DNA replication; gene regulation; gene expression; da.
XX
XX Mykococcus xanthus.
XX
XX US6833447-B1.
PN
XX
XX 21-DEC-2004.
PD
XX
XX 10-JUL-2001; 2001US-00902540.
PP
XX
XX 10-JUL-2000; 2000US-0217883P.
PR
XX
XX (MONS) MONSANTO TECHNOLOGY LLC.
PA

XX Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;
PI
XX
XX WPI; 2005-028716/03.
DR
XX
XX New substantially purified *Myxococcus xanthus* nucleic acid molecule
PT encoding a nitrite reductase, useful for determining gene expression,
PT identifying mutations in a gene of interest, and for constructing
PT mutations in a gene of interest.
XX
XX Example 1; SEQ ID NO 1106; 25pp; English.
XX
XX The invention relates to a substantially purified nucleic acid molecule
CC encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a
CC recombinant DNA construct for expression of a nitrite reductase gene in a
CC plant cell, and a plant cell comprising the recombinant DNA construct.
CC The nucleic acid is useful for determining gene expression, identifying
CC mutations in a gene of interest, and for constructing mutations in a gene
CC of interest. Sequences given in records for SEQ IDs 1-1849 represent a
CC set of 1849 contig and singleton sequences comprising coding sequences,
CC DNA replication elements, promoters and other regulatory elements from
CC the genome of the bacterium *Myxococcus xanthus*. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO
XX
XX Sequence 14638 BP; 2293 A; 4640 C; 5266 G; 2439 T; 0 U; 0 Other;

Query Match	5.1%;	Score 219.8;	DB 14;	Length 14638;
Best Local Similarity	56.5%;	Pred. No. 2.1e-25;		
Matches 429;	Conservative	0;	Mismatches 327;	Indels 3; Gaps 1;
Qy	3009	CGGAGCGTGGCGGTTCAGAGAAATCGTTCCGCTGTGGCATCGACGTACTGCGCGGTGAT	3068	
Db	12645	TTGGAGCGCGCGGTTCAGAGATCATCCGCGCGAGGSCCTCGATGCGCTGGCCGCTCAC	12586	
Qy	3069	CCACCGTGTAGTCGTCCGAGGCGAGAAAGGCCACACGTCGCGCGATGTCGTCCGGTGTGCC	3128	
Db	12585	CCAGCGGCTCTCCGCGCAGCAGCATGGCGACCGCATCCCGATGTGCTCGGCGTCAACC	12526	
Qy	3129	GATGCGGTTGAACACGGAGTGTGGCGGCCAGGTGCGCGCGCGCTCGGGGTCTGCCCGCG	3188	
Db	12525	CACACGCCGAGCGCAACCTGGCGCGGCAGCGGTCTTTGTTCAGCTCCGGGTGTTCACGGAC	12466	
Qy	3189	CGGTGCGTTC--ATGTCCGTCTCCACGAACCCCGGCCACCGCGTTTGACCGGTGATCCC	3245	
Db	12465	GACCGCGCCCGAGTCCGTCCGCGTCCGCGCCGCGCGAGACGTTGACGGAAATCCT	12406	
Qy	3246	CGGTTCCCCAGTTGCTGGCCAGGCGAGCGGTGAGCGTGTTCACCGCACCCCTTGGTGCAT	3305	
Db	12405	CGGGGCCCCAGCTCTTCGCGAGGTACCGGCTGAGCGCTCTCACGCCCGCTTTCATGGT	12346	
Qy	3306	CGGTTATCCGATGGACTCGGGGAAACGCGGCGCGGTTCGGCGACGACGAGATGTTGATGAT	3365	
Db	12345	CGGTTAGCGGGGTGGCCAGGGAAGGTGAAGCGCGCGAGCCCGGAGGAGATGTTGAGAT	12286	
Qy	3366	CGCGCGCGCTCGCGCAGTCGTTTCAGTCCGCTGTGACACAGGAAACACGCGTGCCCGGAC	3425	
Db	12285	GCSCCCCCCATCCGCAGCACCGGCCAACAGCGCTTCGTGAGGAAGAAGTCCCTTGAG	12226	
Qy	3426	GTTTACGGCGGACACGATTCGCTCGAAGACCTCTTCGCTGATCTTCGCTGATCCGTCGAGCC	3485	
Db	12225	GTGGAAGTTTCATCAGCGCGTCGAACCTGGGCGCTCGGTCTCTCCGGAAGCTGACTTGGAT	12166	
Qy	3486	GCTTGAACCGCGGTGTTTACACAGATATGCAACTCGGGCGGCACCTCGAACTCGCCCAT	3545	
Db	12165	GCCCATGCGCGGTGTTTTCACAGGAAGTTCAGGCGCTCGCGCCCGAAGTCAAGGCCCCAG	12106	
Qy	3546	CCGGCGCTCGAACCGCGCGTAGAGCGCGCGCGCTCACCCAGACCGCGAGTTCGGGCCCG	3605	
Db	12105	CTCCGTCCGCAACCGCTTCACAGAGGCGCGAAGCCCGCGGTGTCCCCCAAGTCGAGCGG	12046	
Qy	3606	GATGGGCCAACCGCTGTCCCGCGCTGTCTCCGATGCTTCGACGCTCTCTCCGCGCGCGCGC	3665	

Dn	12045	GAGCGCCACCGCTTGGCCGCTGTGCTTCAATCTGCTTACAGACCGCGCTTCGTC	11986
Qy	3666	CTCGTGTGCGTAGTGGACTGCCACGAGCCCGCTCCGGCGCAGCGCGCGAT	3725
Dn	11985	CGCGCGGTGCGGTACGTACGATGATCCGCTGCCCCGGCGCCAGCTTGAGCGCAT	11926
Qy	3726	ACCGGTCGATGCGCCCGCTTCCCGGTCAACACAGGCG	3764
Dn	11925	GTTCCGCCCATGTCACGGCTTCCCCCGGTGATGAGAGC	11887
RESULT 14			
ID	ACL65960/c		
XX	ACL65960 standard; DNA; 744 BP.		
AC	ACL65960;		
XX			
DT	02-JUN-2005 (first entry)		
DE	M. xanthus gene sequence, seq id 2423.		
XX			
XX	Transgenic plant; DNA replication; gene regulation; gene expression;		
KW	gene; ds.		
XX			
OS	Myxococcus xanthus.		
XX			
FN	US6833447-B1.		
XX			
PD	21-DEC-2004.		
XX			
PF	10-JUL-2001; 2001US-00902540.		
XX			
PR	10-JUL-2000; 2000US-0217883P.		
XX			
PA	(MONS) MONSANTO TECHNOLOGY LLC.		
XX			
PI	Goldman BS, Hinkle GJ, Slater SC, Wiegand RC;		
XX	WPI; 2005-028716/03.		
XX			
PT	New substantially purified Myxococcus xanthus nucleic acid molecule		
PT	encoding a nitrite reductase, useful for determining gene expression,		
PT	identifying mutations in a gene of interest, and for constructing		
PT	mutations in a gene of interest.		
XX			
PS	Example 2; SEQ ID NO 2423; 25pp; English.		
XX			
CC	The invention relates to a substantially purified nucleic acid molecule		
CC	encoding a nitrite reductase of SEQ ID NO. 11926. Further disclosed is a		
CC	recombinant DNA construct for expression of a nitrite reductase gene in a		
CC	plant cell, and a plant cell comprising the recombinant DNA construct.		
CC	The nucleic acid is useful for determining gene expression, identifying		
CC	mutations in a gene of interest, and for constructing mutations in a gene		
CC	of interest. Sequences given in records for SEQ IDs 1850-9691 represent a		
CC	set of about 7842 genes or partial genes from the genome of the bacterium		
CC	Myxococcus xanthus. Note: The sequence data for this patent did not form		
CC	part of the printed specification, but was obtained in electronic format		
CC	directly from USPTO		
XX			
SQ	Sequence 744 BP; 113 A; 235 C; 287 G; 109 T; 0 U; 0 Other;		
Query Match			
Best Local Similarity 5.0%; Score 216; DB 14; Length 744;			
Matches 421; Conservative 0; Mismatches 320; Indels 3; Gaps 1;			
Qy	3024	TCAGAGATCGTTCGCTGTGGCATCGACGCTACTCGCGGTGATCCACCGTGAAGTCGTC	3083
Dn	744	TCAGAGATCATCCCGCCGAGGCTCGATGCGTGGCCCGCTACCCAGCGCTCTCCGG	685
Qy	3084	GGAGCGCAGAAAGGCCACGATCGCGGATGTCGTGGGTCTGCGGATGCGTTGAACAC	3143
Dn	684	CGACAGCAGCATGGCGACCGCATCCCGCTGCTGCTGGGCTGATCCACACACGCGCGCGC	625

RESULT 15			
ID	ADT44601/c		
XX	ADT44601 standard; cDNA; 741 BP.		
AC	ADT44601;		
XX			
DT	02-DEC-2004 (first entry)		
XX			
DE	Bacterial polynucleotide #19352.		
XX			
KW	Recombinant DNA construct; transformed plant; improved plant property;		
KW	cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;		
KW	pathogen tolerance; pest tolerance; plant disease resistance;		
KW	cell cycle pathway modification; plant growth regulator;		
KW	homologous recombination; seed oil yield; protein yield; carbohydrate;		
KW	nitrogen; phosphorus; photosynthesis; lignin; galactomannan;		
XX	bacterial polynucleotide; gene; ss.		
OS	Bacteria.		
XX			
PN	US2003233675-A1.		
XX			
PD	18-DEC-2003.		
XX			
PF	20-FEB-2003; 2003US-00369493.		
XX			
PR	21-FEB-2002; 2002US-0360039P.		
XX			

PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
PI WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 43039; 122bp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 741 BP; 112 A; 235 C; 286 G; 108 T; 0 U; 0 Other;

Query Match 4.9%; Score 213; DB 13; Length 741;
Best Local Similarity 56.4%; Pred. No. 2.7e-24;
Matches 418; Conservative 0; Mismatches 320; Indels 3; Gaps 1;
QY 3027 GAGAAATCGTTCCGCTGTGGCATCGACGTACTGGCCGGGTGATCCACCGTGAGTCGTGCGA 3086
DB 741 GAGCATCATCCGCCCGAGGCGCTCGATGCGCTGGCCCGTCACCCACGCGCTCTCCGGCGA 682
QY 3087 GSCCAGAAAGCCACCACTCGGGGATGTCGTCGGGTCTGCGATGCGGTGTAACACGGA 3146
DB 681 CAGCAGATGGCGACCGCATCCCGATGTCGCGGCTGACCCACACGCGCGGCGCAAC 622
QY 3147 GTTGGCGGCGAGTCCGCGCGCGCTTCGGGGGTCTGCGCGCGCGTTC---ATGTC 3203
DB 621 CTGGGCGGCGCAGCGCTTGTTCAGCTCCGGGTGTCAAGGAGACGCGCGCCCGAAGTC 562
QY 3204 CGTCTCCACGAAACCCGCGGCGCACCGCGTTGACCGGTGATCCCGGTTCCCGCAGTTGCT 3263
DB 561 CGTCCGGGTGCGCGCGCGCGGCGGACGTTGACGAAATCTCCGGGCGCGCCAGCTCCTT 502
QY 3264 GGCACGGCGAGGTGAGGTGTCACCGCGACCCCTTGGTCATCGCGTATCCGATGGACTC 3323
DB 501 CGCGAGGTACCGGGTGAGCGGCTTCCACGCGCGCGCTTCATGGTCGCGGCGGCGGCGC 442
QY 3324 GGGGAAACGCGCGCGGTGCGGCGACAGAGATGTTGATATCCGCGCGCGCTCGCGCAG 3383
DB 441 AGGGAAGGTGAAGCGCGGAGCGCGGAGGAGATGTTGAGGATGCGCGCGCGCGCGCAG 382
QY 3384 TCCTTTTCAGTCCGTGCTGGACAGGAAACAGCGGTGCGCGGACGTTGACGCGCGACCACTCG 3443

Search completed: January 18, 2006, 10:27:13
Job time : 2379 secs

Blank Page (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 18, 2006, 09:45:41 ; Search time 21074 Seconds
(without alignments)
11722.579 Million cell updates/sec

Title: US-10-017-471B-19
Perfect score: 4346
Sequence: 1 gtcagcagcggtcggtt.....gcgtagagcgcggtacc 4346

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_om.*
5: gb_ov.*
6: gb_pac.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_sts.*
11: gb_sy.*
12: gb_un.*
13: gb_vi.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4346	100.0	6727	1	SC0007731
2	4299.8	98.9	290850	1	SC0939127
3	901.2	20.7	12070	1	AB011413
4	563	13.0	210614	1	AB088224
5	527.4	12.1	1203	1	STWAFSAA
6	441.8	10.2	6252	1	SVU24659
7	377.2	8.7	110000	1	BA000030.43
8	347.2	8.0	5495	1	AY228176S2
9	335	7.7	290850	1	SC0939127
10	322.8	7.4	17455	1	AY956334
11	322.6	7.4	16643	1	AY034378
12	316	7.3	39428	1	AF322256
13	314.4	7.2	34869	1	AF324838
14	305.2	7.0	343243	1	BX640414
15	303.8	7.0	348997	1	BX640427
16	303.6	7.0	348706	1	BX640445
17	302.6	7.0	8018	1	AF164960
18	289.8	6.7	92294	1	AY117439

19	280	6.4	2234	1	AB001683
20	279.4	6.4	34644	1	AF080235
21	277.6	6.4	1215	1	AB001608
22	276.8	6.4	300100	1	SC0939123
23	274.2	6.3	88400	6	CQ924550
24	255.4	5.9	7584	6	A81684
25	255.4	5.9	7584	6	A81685
26	255.4	5.9	7584	6	AR633953
27	255.4	5.9	7584	6	AR633954
28	255.4	5.9	7599	1	AF274045
29	255.4	5.9	7600	6	AR437655
30	255.4	5.9	7600	6	AX026699
31	253.6	5.8	110000	1	BA000030.24
32	242.2	5.6	5998	1	AB032524
33	242.2	5.6	110000	1	BA000030.11
34	239.4	5.5	302007	1	SC0939132
35	238	5.5	936	1	AY262284
36	229	5.3	699	6	BD105336
37	229	5.3	1755	1	STWBARA
38	225.4	5.2	32870	1	AF007101
39	219.8	5.1	14638	6	AR619711
40	216	5.0	744	6	AR621022
41	209.6	4.8	2234	1	AB001683
42	209	4.8	2232	1	AY640376
43	209	4.8	2637	1	AF156161
44	208.4	4.8	10480	1	AE008375
45	208.4	4.8	10575	1	AE009238

ALIGNMENTS

RESULT 1

SC0007731
LOCUS Streptomyces coelicolor scbR gene, 6727 bp DNA linear BCT 15-APR-2005
DEFINITION Streptomyces coelicolor scbR gene, scbA gene, ORFs A,B,X & Z.
ACCESSION AJ007731
VERSION gamma-butyrolactone binding protein; scbA gene; scbR gene.
KEYWORDS Streptomyces coelicolor A3(2)
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.

REFERENCE
AUTHORS Takano,E., Chakraborty,R., Nihira,T., Yamada,Y. and Bibb,M.
TITLE Characterisation of scbR, and scbA genes involved in gamma-butyrolactone binding and synthesis in Streptomyces coelicolor
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 6727)
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1998) Takano E., Dept. Genetics, John Innes Institute, Colney lane, Norwich, NR4 7UH, UK

FEATURES
Location/Qualifiers

1..6727
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="M145"
/db_xref="taxon:100226"
37..969
/gene="orfB"
37..969
/gene="orfB"
/codon_start=1
/transl_table=11
/product="histidine kinase"
/protein_id="CAA07625.1"
/db_xref="GI:3425858"
/db_xref="GOA:O86849"
/db_xref="InterPro:IPR000209"
/db_xref="InterPro:IPR003594"
/db_xref="InterPro:IPR011712"

```
/db_xref="UniProt/TREMBL:O86849"
/translation="MGTPSPLENRRGGLLIPCSLDPRVSRRTVERHFCBPTPARHD
LAGSHRGGRFTETLODGLSRFRGARQTTGTAHPGLPAGRDPAAGPVRTDRGRDD
HWAEARRAREDELGTALSTAAHHIELHAAETGGTHLDAARHSLEAMAVARRL
TGELOAQTVPPLAQAVGEFAATTPGCTEVRITRTGGDELLSDVCRRELFVLAVERAL
NHAFHAAHADVTTLRTRWAAHAGIVDDGVGFDADAVLVFGHRAFGRLSRMTDRIED
VGMRLIVSGPAGGTHIDVHLPLPRKVKSTAPRI"
complement(1000. .1854)
/gene="orfA"
complement(1000. .1854)
/gene="orfA"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAA07626.1"
/db_xref="GI:3425859"
/db_xref="UniProt/TREMBL:O86850"
/translation="MWLVGTGVGVGVVAHLLARTGVRVLARRPVRTVTGPNVEV
CAGYADRPGLDRALNGVRAAPLVNNSATEPDDERFAAAREAGVRLVKLSMLAVGE
PGADDFITRRQRENERAVRESGLDWTFLKARTFMSNTLSWAPAIRSDGVVRALYGTSP
VACVDPRDIAEVAVRALTRPGHGRVALSGFEAITAQQTARLSEVLGRLSRPEELG
LEAAREOMLRYPVPVADAFLESAERQAGAKAQVVTREVTGRPARPFIWAADHA
BAFGRSQG"
complement(1970. .2914)
/gene="scbA"
complement(1970. .2914)
/gene="scbA"
/function="involved in gamma-butyrolactone synthesis"
/codon_start=1
/transl_table=11
/product="ScbA protein"
/protein_id="CAA07627.1"
/db_xref="GI:3425860"
/db_xref="UniProt/TREMBL:O86851"
/translation="WPEAVLINSASDANSIEQTALPVPMLVHRTRVQDAPVPSWIP
KGRDFSTAVLPHDHPFPAVHGDRDHPLLIATLRQAAAMLVPHAGVGVPGVHFLM
ATLDYTHLDHLDGHPFVAEIEVACSQLPRGGQVQGVQVDWAVRAGRLAATGTA
TTRFTSPQVYRMGRGFATFASVPGTAPVPAARGRTRDEDVVLSSQQDTWRLRV
DTSHTPLQRPNDRHVPGLLEAAQAAACLVTPGAPFVPSIGSTRFVRYAEFDSFCWI
QATVRPGPAAGLTTRVTGHQDGLVFLTLTGPAFSG"
3032. .3679
/gene="scbR"
3032. .3679
/gene="scbR"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="gamma-butyrolactone binding protein"
/protein_id="CAA07628.1"
/db_xref="GI:3425861"
/db_xref="GOA:O86852"
/db_xref="UniProt/TREMBL:O86851"
/db_xref="UniProt/TREMBL:O86852"
/translation="WAKQDRALTRQTILDAAAQVFEKQVQAATITELIKVAGTKG
ALYFFHQSKELALVDAQPEPPQVPEQRLQELIDMGMLFCHRLTNVVRAGVR
LSMDQAHGLDRGRGFRWHETLLKLNAKENELLPHVVTTFSDIYNVTFAGIQV
VSQTVSDYQDLQLEHRYALLQKHLLPAIAVPSVLAALDLSEERGARIAELAPTGD"
complement(3795. .4529)
/gene="orfX"
complement(3795. .4529)
/gene="orfX"
/codon_start=1
/transl_table=11
/product="3-ketoacyl-ACP/CoA reductase"
/protein_id="CAA07629.1"
/db_xref="GI:3425862"
/db_xref="GOA:O86853"
/db_xref="UniProt/TREMBL:O86853"
/db_xref="UniProt/TREMBL:O86853"
/translation="WTGSGRGIGRIALRLAADGALVAVHYGSSFAAARETVETIRSS
GGQALAIRAELGVGDAALYAAPDAGMGFCVFPFEDLILVNNAGVSGSGRITVETEE
```

gene

CDS

ORIGIN

```
Query Match      100.0%; Score 4346; DB 1; Length 6727;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1  GTCACGACGGCGTTCGGGTTTCGACGCGACGGGTACTCGTTCCGCGCACCGGSCACCG 60
DB 772 GTCACGACGGCGTTCGGGTTTCGACGCGACGGGTACTCGTTCCGCGCACCGGSCACCG 831

QY 61  GGTCTGGCTCGATGACCGACCGCATCGAGGACGTCGGCTCGCGGCTCCTGATAGTGAGC 120
DB 832 GGTCTGGCTCGATGACCGACCGCATCGAGGACGTCGGCTCGCGGCTCCTGATAGTGAGC 891

QY 121 GGGCCCGCGCGGACGACATCGAGTCATCTCCACATGCGGCCCGCCCGGAAAGTGAGC 180
DB 892 GGGCCCGCGCGGACGACATCGAGTCATCTCCACATGCGGCCCGCCCGGAAAGTGAGC 951

QY 181 ACCGACCGCGGAGTGACGCGCATGGGAGGGCCACGTCGCGGACGGATCACCCCTGGCT 240
DB 952 ACCGACCGCGGAGTGACGCGCATGGGAGGGCCACGTCGCGGACGGATCACCCCTGGCT 1011

QY 241 TCGGCGGAAGCTTTCGGCTGGTTCGCGCCCGCCAGATGCGGAACGCGCTTGGCGGGCGGCG 300
DB 1012 TCGGCGGAAGCTTTCGGCTGGTTCGCGCCCGCCAGATGCGGAACGCGCTTGGCGGGCGGCG 1071

QY 301 CGTCACTTCCGCGACGGTCGCGACGACCTGGCGGCTTGGCCCGCCCGCTGCGCGCTCGGC 360
DB 1072 CGTCACTTCCGCGACGGTCGCGACGACCTGGCGGCTTGGCCCGCCCGCTGCGCGCTCGGC 1131

QY 361 GCTCTCCAGGAACGGTCGCGGACGGGCTCGGATCTTCCGAGCATCTGCTGCGCGC 420
DB 1132 GCTCTCCAGGAACGGTCGCGGACGGGCTCGGATCTTCCGAGCATCTGCTGCGCGC 1191

QY 421 CGCTTCCAGGCCCGAGCTCCTCGAAACGAGTGACCGCCCGCCAGCACCTCGGAGAGCCGCGC 480
DB 1192 CGCTTCCAGGCCCGAGCTCCTCGAAACGAGTGACCGCCCGCCAGCACCTCGGAGAGCCGCGC 1251

QY 481 CGTCTGCTCGCTGGCGGTGATCGCTCGGGCCCGGACAGCGGTACGCGCGTCCCTCGTG 540
DB 1252 CGTCTGCTCGCTGGCGGTGATCGCTCGGGCCCGGACAGCGGTACGCGCGTCCCTCGTG 1311

QY 541 GCGGGCCCGGTGAGTGCCTTGAACCGCCACTTCCGCGATGTCGCGCGATCGACGACG 600
DB 1312 GCGGGCCCGGTGAGTGCCTTGAACCGCCACTTCCGCGATGTCGCGCGATCGACGACG 1371

QY 601 AACCGGGACGTCGCGGTACAGCGCGCGACCAACGCGCTCGGACCGGATGGCGGGCGCCA 660
```

Db 1372 AACCGGGACGTCGCTACAGCGCGCGGACACAGCCGTCGACCGGATGGCGGGCGCCCA 1431
Qy 661 GGACACGCTGTGGACATGAAGGTCTCTGGCCCGCAGGAAGTCCAGTCTAGCCCGGATC 720
Db 1432 GGACACGCTGTGGACATGAAGGTCTCTGGCCCGCAGGAAGTCCAGTCTAGCCCGGATC 1491
Qy 721 GGGTACGGCCGCTCGTCTCGGCTGCGCGCGCTGATGAAGTCTGCGCGCGCGTTC 780
Db 1492 GCGTACGGCCGCTCGTCTCGGCTGCGCGCGCTGATGAAGTCTGCGCGCGCGTTC 1551
Qy 781 CCCACCGCGACATGAAGTCTCAACAGGTGCGCGACCGCGCTCGCGCGCGCGCGC 840
Db 1552 CCCACCGCGACATGAAGTCTCAACAGGTGCGCGACCGCGCTCGCGCGCGCGCGC 1611
Qy 841 CGCGAAACGCTCGTCTCGGCTCGGTCAGTCTGTCGACAGGAACCGCGCGCGCAC 900
Db 1612 CGCGAAACGCTCGTCTCGGCTCGGTCAGTCTGTCGACAGGAACCGCGCGCGCAC 1671
Qy 901 CCGTGTGAGGCGCGCTCAGGCGCGGCGGTGCGGCTACTCGCGCGCGACACTCGAC 960
Db 1672 CCGTGTGAGGCGCGCTCAGGCGCGGCGGTGCGGCTACTCGCGCGCGACACTCGAC 1731
Qy 961 GTTCCGGCGCGTGAAGTCAACCGTCTCGGCGCGCGCGGAGACTCTGAACGGGACCGGT 1020
Db 1732 GTTCCGGCGCGTGAAGTCAACCGTCTCGGCGCGCGCGGAGACTCTGAACGGGACCGGT 1791
Qy 1021 CCGGCGCAGCAGTGGCGACCTGACGGCGGACCAACCGGTGACCGCGGTCAAGAAT 1080
Db 1792 CCGGCGCAGCAGTGGCGACCTGACGGCGGACCAACCGGTGACCGCGGTCAAGAAT 1851
Qy 1081 CACTCGGCGCTCTCTCGGCGAGCAGGCGAGGGGCGCTCCGAAATATATAGAGGGA 1140
Db 1852 CACTCGGCGCTCTCTCGGCGAGCAGGCGAGGGGCGCTCCGAAATATATAGAGGGA 1911
Qy 1141 AGGCGAGGATCTGCCCCCGGGCGGAAACCGGCGATGTTGCGCCCCCGGGCGGTTC 1200
Db 1912 AGGCGAGGATCTGCCCCCGGGCGGAAACCGGCGATGTTGCGCCCCCGGGCGGTTC 1971
Qy 1201 AGCGGAGAAACCGGGCGCGGACAGCGTGTGAGGAAGCAGGCTGCGCTCTGATGCC 1260
Db 1972 AGCGGAGAAACCGGGCGCGGACAGCGTGTGAGGAAGCAGGCTGCGCTCTGATGCC 2031
Qy 1261 CCGTGAACCGGACAGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320
Db 2032 CCGTGAACCGGACAGTGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2091
Qy 1321 AGCAGGGCTGTGAACTCCGGGTACCGGACGAAACCGGCTGCGCGGATCGACGACGGA 1380
Db 2092 AGCAGGGCTGTGAACTCCGGGTACCGGACGAAACCGGCTGCGCGGATCGACGACGGA 2151
Qy 1381 AGGGCGCGCGGACCGGTACAGGCGACCGCGCTGCGCGCTCGAGCAGCAGCATGC 1440
Db 2152 AGGGCGCGCGGACCGGTACAGGCGACCGCGCTGCGCGCTCGAGCAGCAGCATGC 2211
Qy 1441 CCGGTACGCTGTGTTGGGGCGGTGAAGAGGGTGGGTGATGTTGTCACCGCGAGTC 1500
Db 2212 CCGGTACGCTGTGTTGGGGCGGTGAAGAGGGTGGGTGATGTTGTCACCGCGAGTC 2271
Qy 1501 GCGAGTGTCTGCTGCGAACTCGCGGACAGGACGACGTCCTGCTGCGGGGTGCGACCGG 1560
Db 2272 GCGAGTGTCTGCTGCGAACTCGCGGACAGGACGACGTCCTGCTGCGGGGTGCGACCGG 2331
Qy 1561 CCGCGCGCGCGGACCGGCGCGGTCCCGGGACACGATCGGTGGGAGTTCGCGAAGTTCG 1620
Db 2332 CCGCGCGCGCGGACCGGCGCGGTCCCGGGACACGATCGGTGGGAGTTCGCGAAGTTCG 2391
Qy 1621 CCGCGCATCCCGGTGATCTGAGGACTGTGAAGCGCTGTGAAGCGTCCCGGTGGCAG 1680
Db 2392 CCGCGCATCCCGGTGATCTGAGGACTGTGAAGCGCTGTGAAGCGTCCCGGTGGCAG 2451
Qy 1681 CCGAGCGTCCGGCGCGGCGACGCGCGCGTCAACCTGCTCTGATCGGCGTCCCGCGCG 1740
Db 2452 CCGAGCGTCCGGCGCGGCGACGCGCGCGTCAACCTGCTCTGATCGGCGTCCCGCGCGCG 2511

Qy 1741 GGAATTCAGCTGGGAACAGGCCACTTCCACTTCAGCTCCGACCTCGCCGACCGC 1800
Db 2512 GGAATTCAGCTGGGAACAGGCCACTTCCACTTCAGCTCCGACCTCGCCGACCGC 2571
Qy 1801 CGAGTGTGTCCAGGTGGCAGGTGTAGTCCAGCGTGGCCATCAGGAAGTGTAGCCACCG 1860
Db 2572 CGAGTGTGTCCAGGTGGCAGGTGTAGTCCAGCGTGGCCATCAGGAAGTGTAGCCACCG 2631
Qy 1861 GCAAGCGTGTAGCCGGCTGGGAAGAGAGCATCGCCGCTGACGCGAGGTCTCGGCGATCA 1920
Db 2632 GCAAGCGTGTAGCCGGCTGGGAAGAGAGCATCGCCGCTGACGCGAGGTCTCGGCGATCA 2691
Qy 1921 GCAGCGGATCTGTGCGGTCCCGTGGACCGGTGCGAAGAACGGGTGTGCTGGGCGAGA 1980
Db 2692 GCAGCGGATCTGTGCGGTCCCGTGGACCGGTGCGAAGAACGGGTGTGCTGGGCGAGA 2751
Qy 1981 CGGCGGTGACCGAGAAACCGGTCAACCGCTTCGCTGATCCAGCTGACCGGGAACGGTCT 2040
Db 2752 CGGCGGTGACCGAGAAACCGGTCAACCGCTTCGCTGATCCAGCTGACCGGGAACGGTCT 2811
Qy 2041 GCACCTCGTCCGGTGGCAAGCGCATCGGAACCGGCAATGCGGTGTTGTCGATCGAGT 2100
Db 2812 GCACCTCGTCCGGTGGCAAGCGCATCGGAACCGGCAATGCGGTGTTGTCGATCGAGT 2871
Qy 2101 TGGCATCGGACGCGAATTTGATCAAACTACTGCTTCGGGCAATGGTCCCCCAGGAAT 2160
Db 2872 TGGCATCGGACGCGAATTTGATCAAACTACTGCTTCGGGCAATGGTCCCCCAGGAAT 2931
Qy 2161 CATGTGATCGCGAGCTGTTCTGTATGCGCGAACTTAAGATACAGACTGAGCGGTGTTT 2220
Db 2932 CATGTGATCGCGAGCTGTTCTGTATGCGCGAACTTAAGATACAGACTGAGCGGTGTTT 2991
Qy 2221 TTCTATCTTCCGGGGGAGACATGAACAAAGAGGACGAGCATGGCAAGAGCAGGAC 2280
Db 2992 TTCTATCTTCCGGGGGAGACATGAACAAAGAGGACGAGCATGGCAAGAGCAGGAC 3051
Qy 2281 GATCCGACCGGGCAGACGATCTTGGAACCGCGGGCGAGGTCTTCGAGAAAGCAGGCTA 2340
Db 3052 GATCCGACCGGGCAGACGATCTTGGAACCGCGGGCGAGGTCTTCGAGAAAGCAGGCTA 3111
Qy 2341 CCAAGCTCCAGATCAACGAGATCTCAAGGTGCGCGGGTGACCAAGGAGCGCTCTTA 2400
Db 3112 CCAAGCTCCAGATCAACGAGATCTCAAGGTGCGCGGGTGACCAAGGAGCGCTCTTA 3171
Qy 2401 CTTTCACTTCCAGTCCAAAGGAAGAACTGGCGCTGGCGGTTCGAGCGCCAGGAACCA 2460
Db 3172 CTTTCACTTCCAGTCCAAAGGAAGAACTGGCGCTGGCGGTTCGAGCGCCAGGAACCA 3231
Qy 2461 ACAGGCGGTTCGGGAGCAACCCCTCGGCTGCAAGAACTCATCTGACATGGGCAATGTT 2520
Db 3232 ACAGGCGGTTCGGGAGCAACCCCTCGGCTGCAAGAACTCATCTGACATGGGCAATGTT 3291
Qy 2521 CTGTCAACCGTTCGCGACGAACTGTGGCCCGGCGCGCGCTCTCCATGGAACCA 2580
Db 3292 CTGTCAACCGTTCGCGACGAACTGTGGCCCGGCGCGCGCTCTCCATGGAACCA 3351
Qy 2581 GCAGCGCACGCTCTCGATCGCCGAGGACCTTCCTGCTGCTGGCAGAGACACTCTCTGAA 2640
Db 3352 GCAGCGCACGCTCTCGATCGCCGAGGACCTTCCTGCTGCTGGCAGAGACACTCTCTGAA 3411
Qy 2641 GCTGCTGAACCAAGGCCAAGAGGAAAGGAGTGTGCTGCCCATGTGGTCAACCGGACTC 2700
Db 3412 GCTGCTGAACCAAGGCCAAGAGGAAAGGAGTGTGCTGCCCATGTGGTCAACCGGACTC 3471
Qy 2701 GCGCGATCTCTACGTGGGCAAGTTCGCGGGGATACAGGTGTGCTCCAGACGCTCAGCGA 2760
Db 3472 GCGCGATCTCTACGTGGGCAAGTTCGCGGGGATACAGGTGTGCTCCAGACGCTCAGCGA 3531
Qy 2761 CTACAGGACCTTCGAACACCGCTACCGGTGCTGCGAAGGACATCTCTGCGCGCCATCGC 2820
Db 3532 CTACAGGACCTTCGAACACCGCTACCGGTGCTGCGAAGGACATCTCTGCGCGCCATCGC 3591

Qy	2821	GGTTCCCTCGTGTGGCGCGCTCGATCTCTCCGAGGAGCGCGAGCAGCCCTCGGGC	2880
Db	3592	GGTTCCCTCGTGTGGCGCGCTCGATCTCTCCGAGGAGCGCGAGCAGCCCTCGGGC	3651
Qy	2881	CGAATGCGACCGACCGGGGAAGGACTGACCGCCGGAAGCGCCGACCGGATACCGAACCG	2940
Db	3652	CGAATGCGACCGACCGGGGAAGGACTGACCGCCGGAAGCGCCGACCGGATACCGAACCG	3711
Qy	2941	CCGTGCCGAGCGCGCGACCGGGCGCGCTACGAGGCGCGGGCGGGCGCGGTAGGTCTGC	3000
Db	3712	CCGTGCCGAGCGCGCGACCGGGCGCGCTACGAGGCGCGGGCGGGCGGGTAGGTCTGC	3771
Qy	3001	CCTCGGTACCGAAGCGTGGCGGGTCAAGAGATCGTTCCGCTGTGGCATCGACGTACTGG	3060
Db	3772	CCTCGGTACCGAAGCGTGGCGGGTCAAGAGATCGTTCCGCTGTGGCATCGACGTACTGG	3831
Qy	3061	CCGGTGTATCCACCGGTAGTCTCGAGGCGCAGAAAGGCCACCAAGTTCGGCGATGTCGTGC	3120
Db	3832	CCGGTGTATCCACCGGTAGTCTCGAGGCGCAGAAAGGCCACCAAGTTCGTGTGC	3891
Qy	3121	GGTTGCGCATGCGGTTGAAACAGGAGTTGGCGGCGAGTGCAGCGCGCGCTCGGGGGTC	3180
Db	3892	GGTTGCGCATGCGGTTGAAACAGGAGTTGGCGGCGAGTGCAGCGCGCGCTCGGGGGTC	3951
Qy	3181	TGCGCGCGCGGTGGTTCATCTCGTCTCCACGAAACCCGCGGCGCACCGCGTTGACCGTG	3240
Db	3952	TGCGCGCGCGGTGGTTCATCTCGTCTCCACGAAACCCGCGGCGCACCGCGTTGACCGTG	4011
Qy	3241	ATCCCCCGTTCCCCCAGTTGCTTGCCACAGGCGAGCGTGAAGCGTTCACCGCACCCCTTG	3300
Db	4012	ATCCCCCGTTCCCCCAGTTGCTTGCCACAGGCGAGCGTGAAGCGTTCACCGCACCCCTTG	4071
Qy	3301	GTCATCGCGTATCCGATGGAATCGGGGAAACGCGCGCGGGTTCGGGCGACAGATGTTG	3360
Db	4072	GTCATCGCGTATCCGATGGAATCGGGGAAACGCGCGCGGGTTCGGGCGACAGATGTTG	4131
Qy	3361	ATGATCCGCGCGCGTTCGGCAGTCTGTTTCAGTCCGTTCTGACCGAGAACAGCGGTGCC	3420
Db	4132	ATGATCCGCGCGCGTTCGGCAGTCTGTTTCAGTCCGTTCTGACCGAGAACAGCGGTGCC	4191
Qy	3421	CGGAGTTGACGGGACAGTCCGTCGGAAGACCTCTCGGTGACTTCCGTTGATCCGTCCC	3480
Db	4192	CGGAGTTGACGGGACAGTCCGTCGGAAGACCTCTCGGTGACTTCCGTTGATCCGTCCC	4251
Qy	3481	GAGCGCTGACCGCGCGTGTTCACAGAGATGCGAACTCGGGCGGCACTTCGAACTCG	3540
Db	4252	GAGCGCTGACCGCGCGTGTTCACAGAGATGCGAACTCGGGCGGCACTTCGAACTCG	4311
Qy	3541	CCCATCCGGCGTGAACCGCGCGTAGAGCGGGCGCGGTCAACACGACCGGAGTTTCG	3600
Db	4312	CCCATCCGGCGTGAACCGCGCGTAGAGCGGGCGCGGTCAACACGACCGGAGTTTCG	4371
Qy	3601	GCCGGATGGCCCAACGCTGTCCGCGCTGTCCCGATGGTCTCGACGGTCTCTCGCGCC	3660
Db	4372	GCCGGATGGCCCAACGCTGTCCGCGCTGTCCCGATGGTCTCGACGGTCTCTCGCGCC	4431
Qy	3661	GCCGCTCGCTGCTCCGCTAGTGACTGCGACGAGCGCCCGTTCGGCGGCGAGCGCAGG	3720
Db	4432	GCCGCTCGCTGCTCCGCTAGTGACTGCGACGAGCGCCCGTTCGGCGGCGAGCGCAGG	4491
Qy	3721	GGGATACCGGTCCGATGCGCCGGCTTCCCGGTTCAACAGGGCGGTCTTGCCCTCCAGC	3780
Db	4492	GGGATACCGGTCCGATGCGCCGGCTTCCCGGTTCAACAGGGCGGTCTTGCCCTCCAGC	4551
Qy	3781	GGTCTTCCATACCTCGTCCCATGTGACGATATCAGCCCGCGCTCGGTGAGCGACCC	3840
Db	4552	GGTCTTCCATACCTCGTCCCATGTGACGATATCAGCCCGCGCTCGGTGAGCGACCC	4611
Qy	3841	ATGGCGCGCGTTCGGCGGTTCGAAATCGAGCGGTTCACAGCTTACCTGTGACCGCGTCAGACG	3900
Db	4612	ATGGCGCGCGTTCGGCGGTTCGAAATCGAGCGGTTCACAGCTTACCTGTGACCGCGTCAGACG	4671
Qy	3901	GGGCGCGGATGCGCCCGGTTGGACCGCTGGGGCCAGATCGGGCGGCGCGACGGGGAAACCG	3960

Db	4672	GGGCGCGGATGCGCCCGGTTGGACGGCTGGGGCGCAGATCGGGCGCGCGCACGGGGAAACCG	4731
Qy	3961	GCGCGCGTCAAGGGTCAAGGGTTCGGCGGACCGCCAGGCGCGGTCAAGGCGACCGACCGGA	4020
Db	4732	GCGCGCGTCAAGGGTCAAGGGTTCGGCGGACCGCCAGGCGCGGTCAAGGCGACCGACCGGA	4791
Qy	4021	TCGAGTTCGGCGGTGCCACCGCGCCACAGTCTTCGCGGGCCAGCTCCGACTCCGTAGCGCG	4080
Db	4792	TCGAGTTCGGCGGTGCCACCGCGCCACAGTCTTCGCGGGCCAGCTCCGACTCCGTAGCGCG	4851
Qy	4081	TACAGAGCCCCGCTTCGGCCGAGTCTGAGCTGAGCGTGGCGCGCGCGGTGGGTGAGCGCG	4140
Db	4852	TACAGAGCCCCGCTTCGGCCGAGTCTGAGCTGAGCGTGGCGCGCGCGGTGGGTGAGCGCG	4911
Qy	4141	TTGCGCCAGGGGCGGAAGCGGGAGGTTCGGCGCGCGAGCATCATGGGGCGGGCGCGGTTCG	4200
Db	4912	TTGCGCCAGGGGCGGAAGCGGGAGGTTCGGCGCGCGAGCATCATGGGGCGGGCGCGGTTCG	4971
Qy	4201	AAACGGCGCGCGCGGTTCAGGGCTCTTCAGAGGACGTCTAGACCGCCCAACCCCGCCC	4260
Db	4972	AAACGGCGCGCGCGGTTCAGGGCTCTTCAGAGGACGTCTAGACCGCCCAACCCCGCCC	5031
Qy	4261	TGCGCGCAGGCGGCGACCGCGCCCGCGCAGCTTCGCGCGTTCGCGCGCGCGGCGAGCGG	4320
Db	5032	TGCGCGCAGGCGGCGACCGCGCCCGCGCAGCTTCGCGCGTTCGCGCGCGCGGCGAGCGG	5091
Qy	4321	AGCGACCGGTAGAGCGCGCGGTACCC 4346	
Db	5092	AGCGACCGGTAGAGCGCGCGGTACCC 5117	
RESULT 2			
SCO939127/c			
LOCUS			
DEFINITION			
SCO939127 AL023496 AL023861 AL031155 AL031182 AL031232 AL0315161			
ACCESSION			
AL035205 AL035206 AL132824 AL512902 AL590982 AL591083 AL591084			
AL645882			
VERSION			
AL939127.1 GI:24429552			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kleser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornaby, T., Howarth, S., Huang, C.H., Kleser, T., Larke, D., Murphy, L., Oliver, K., O'Neil, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorek, A., Woodward, J., Barrell, B.G., Parkhill, J. and Hopwood, D.A.			
Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)			
Nature 417 (6885), 141-147 (2002)			
12000953			
2 (bases 1 to 290850)			
Bentley, S.D.			
Direct Submission			
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk			
On or before Oct 30, 2002 this sequence version replaced			
gi:20520661, gi:20520749, gi:20520754, gi:20520755, gi:20520688, gi:20520689, gi:20520815, gi:20520683, gi:20520917, gi:20520686, gi:20520864.			
TITLE			
coelicolor A3(2)			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
1. .290850			
/organism="Streptomyces coelicolor A3(2)"			
/mol_type="genomic DNA"			

```

/strain="A3(2)"
/db_xref="taxon:100226"
97. .1101
/gene="SCO6257"
/notes="synonym: SCAH10.22"
97. .1101
/gene="SCO6257"
/notes="SCAH10.22, probable ABC transport system sugar binding lipoprotein, len: 334 aa; similar to TR:CA841563 (EMBL:AL049727) Streptomyces coelicolor putative secreted solute binding protein, 337 aa; fasta scores: opt: 1003 z-score: 1130.9 E(): 0; 49.8% identity in 325 aa overlap, to SW:RBSB_DACSU (EMBL:Z92953) Bacillus subtilis D-ribose-binding protein precursor RbsB, 305 aa; fasta scores: opt: 206 z-score: 238.7 E(): 6.3e-06; 23.0% identity in 235 aa overlap and to TR:CA841563 (EMBL:SC981) Streptomyces coelicolor SC9B1.17c, 337 aa; fasta scores: opt: 1003 z-score: 1045.2 E(): 0; 49.8% identity in 325 aa overlap. Contains match to Pfam entry PF00532 Peripla_BP_like, Periplasmic binding proteins and LacI family and a match situated in the correct position to Prosite entry PS00013 Prokaryotic membrane lipoprotein lipid attachment site. Contains also possible N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative ABC transport system sugar binding lipoprotein"
/protein_id="CAB60176.1"
/db_xref="GI:6273662"
/db_xref="InterPro:IPR001761"
/db_xref="UniProt/TREMBL:Q9RKT5"
/translation="MARFTWVGIAGALGALSVAQVNDINFLYSHDAEQOALVNAA ADTPRTFAMITHSGDGTWDIVQSGAEQAAVDNINFLYSHDAEQOALVNAA IDKVDGLIYTLAKDPAKSAARAHKAGIPVITVNSGSESKERFGLTHVGDDETTA GRAVEBELNREQAVLGHQGNVGHQSCDGVKEFTDKVQRLYNGTSMPEVQS ATEAKLQDKSDVAVTLCVHPYADTAVKAKQAGSKAEITFDLNKAKVAAGLADGTLG FAVDQOPLYQGEAVDILLKYNDVLVGGSPVLTGPQIITKDDAALADYTKRGT"
133. .165
/misc_feature
/gene="SCO6257"
/notes="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
241. .966
/misc_feature
/gene="SCO6257"
/notes="Pfam match to entry PF00532 Peripla_BP_like, periplasmic binding proteins and LacI family., score 13.80, E-value 2.1e-06"
1098. .2138
/gene="SCO6258"
/notes="synonym: SCAH10.23"
1098. .2138
/gene="SCO6258"
/notes="SCAH10.23, probable ABC transport system sugar permease, len: 346 aa; similar to TR:O68120 (EMBL:AF010496) Rhodospirillum rubrum capsule protein RbsC, 345 aa; fasta scores: opt: 445 z-score: 508.0 E(): 6.4e-21; 30.5% identity in 351 aa overlap and to SW:RBSB_ECOLI (EMBL:L10328) Escherichia coli ribose transport system permease protein RbsC, len: 321 aa; fasta scores: opt: 286 z-score: 329.9 E(): 5.3e-11; 30.1% identity in 329 aa overlap. Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative ABC transport system sugar permease"
/protein_id="CAB60177.1"
/db_xref="GI:6273663"
/db_xref="GOA:Q9RKT4"
/db_xref="InterPro:IPR001851"
/db_xref="UniProt/TREMBL:Q9RKT4"
/translation="MNATQVKAADRIQLTSTRIKLLGRPELGSVVGAIVAVFFAFF ADSFLHAASLSTLVYAASTIGINAVPALLMIGDFDLSAGVNVVTSSALVSMFSYQM
TANVGVVSVLLVTLTAIGAFNGFMLTRTKLPSPFITLTGTFMLTGLNLGPTKLVGDT
VSTKSIADMEGFPSPADQVFASTITIGGVGFKVTLMLWALVAVASMLLRTPRAGNWIF
AVGNGDAARAVGVFPVAKTIGLTMGVGFGAWISQHLHLSFYDVSQSEGVGNELIYI
IAAVIGGLTGGYGSVAGSAGVAFICGTSGKIVFABWNPDPWFKFLGAMLLATLL
NAWVRKRAEATK"
2124. .2127
/gene="SCO6258"
2135. .2926
/gene="SCO6259"
/notes="synonym: SCAH10.24"
2135. .2926
/gene="SCO6259"
/notes="SCAH10.24, probable ABC sugar transport ATP binding protein, len: 263 aa; similar to TR:O68121 (EMBL:AF010496) Rhodospirillum rubrum capsule protein RbsC, 305 aa; fasta scores: opt: 612 z-score: 706.1 E(): 5.9e-32; 45.6% identity in 217 aa overlap, to SW:RBSA_ECOLI (EMBL:M13169) Escherichia coli, ribose transport ATP-binding protein RbsA, 501 aa; fasta scores: opt: 509 z-score: 585.6 E(): 3e-25; 31.5% identity in 254 aa overlap and to TR:CAAL5787 (EMBL:SC787) Streptomyces coelicolor St7B7.07, 260 aa; fasta scores: opt: 592 z-score: 521.2 E(): 1e-23; 43.1% identity in 246 aa overlap. Contains Pfam match to entry PF00005 ABC tran, ABC transporter and Prosite PS00017 ATP/GTP-binding site motif A (P-loop) and PS00211 ABC transporters family signature"
/codon_start=1
/transl_table=11
/product="probable ABC sugar transport ATP binding protein"
/protein_id="CAB60178.1"
/db_xref="GI:6273664"
/db_xref="GOA:Q9RKT3"
/db_xref="InterPro:IPR003439"
/db_xref="UniProt/TREMBL:Q9RKT3"
/db_xref="UniProt/TREMBL:Q9RKT3"
/translation="MTRNEDTALVELSGVNGVNRALGVSLVHAGETLCVLGD NGAGSTLIKIAGLHQHDDGTLSDGETRLSSPREADRGRTVYQDLAVPLMPV WRNFFLSEPRKGVAPFKRMDVDMRRTTHAELRMGIDLDRVDQPIGLSGERGOCV AIARAVYFGAKVLVDLEPTALGVQSGVLYKVAARDQGLGVLLITHPHAYLVG DRFVLLKRGTMVANQTRDBVTLDLTQMQAGTDLDLRHELRG"
2240. .2815
/gene="SCO6259"
/notes="Pfam match to entry PF00005 ABC tran, ABC transporter, score 168.70, E-value 9.6e-47"
2261. .2284
/gene="SCO6259"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
2588. .2632
/gene="SCO6259"
/notes="PS00211 ABC transporters family signature"
3042. .4190
/gene="SCO6260"
/notes="synonym: SCAH10.25"
3042. .4190
/gene="SCO6260"
/notes="SCAH10.25, possible sugar kinase, len: 382 aa; similar to SW:GLK_STRCO (EMBL:X65932) Streptomyces coelicolor glucokinase (EC 2.7.1.2) (Glucose kinase) Glk, 317 aa; fasta scores: opt: 1022 z-score: 1142.9 E(): 0; 46.6% identity in 311 aa overlap and to TR:CA851974 (EMBL:SC610) Streptomyces coelicolor SC610.20c, 317 aa; fasta scores: opt: 1022 z-score: 1035.4 E(): 0; 46.6% identity in 311 aa overlap. Contains a match to Pfam entry PF00480 ROK_ROK family and Prosite entry PS01125 ROK family signature"
/codon_start=1
/transl_table=11
/product="putative sugar kinase"
/protein_id="CAB60179.1"
/db_xref="GI:6273665"
/db_xref="GOA:Q9RKT2"

```

```
/db_xref="InterPro:IPR000600"
/db_xref="InterPro:IPR004654"
/db_xref="UniProt/TREMBL:Q9RKT2"
translation="MSTVRDPTAFIGRRAPVLRVTGTRRRSHLTAPRVPTVGIDIG
GTVKAGVVDADGNLILKLRITETPKSKPRKVEDTIVELVLDLSRDHVDHVGIGAA
GWDADNRNVLFAPHLSWRNPLRDIAGRLAVPLVDNDANTAAWAEWFGAGRGD
HLVMTLTGTGGIIGDGOVKRGKYGVAGFEHMQVVPGGHRCPCPNRGCGWEYSSG
NALVREARELAASDPVAYGLIEHVKSIGDITGPMTELAREGDAMCVRELLQIDQW
LGVGIANLAALADPSCFVIGGVSAADDLLIGADAPFKROLTGRTGRTPEARIVRQL
GPEAGWGAADLSLVARFRFRARRRVERRYERYAEARRRESRL"
misc_feature
3159..3704
/gene="SC06260"
/note="Pfam match to entry PF00480 ROK, ROK family, score
215.50, E-value 5.4e-63"
3540..3623

misc_feature

Query Match          98.9%;   Score 4299.8;   DB 1;   Length 290850;
Best Local Similarity 99.8%;   Pred. No. 0;
Matches 4337; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

QY 1 CTCGACGACGGCGTCCGGTTTCGACGCCGACGCGGTACTCGTTCCGGGCCACCGGGCACCG 60
DB 14234 GTCGACGACGGCGTCCGGTTTCGACGCCGACGCGGTACTCGTTCCGGGCCACCGGGCACCG 14175
QY 61 GGTCTGCGCTCGATGACGACCGCATCGACGCTCGACGCTCGGCTCGCGCTCCTGATAGTGAGC 120
DB 14174 GGTCTGCGCTCGATGACGACCGCATCGACGCTCGGCGGCGGCTCCTGATAGTGAGC 14115
QY 121 GGCCCCCGCGGGCAGCAGCATCGACGCTCATCTCCCACTGCGCCCCCGGAAAGTGTAGC 180
DB 14114 GGCCCCCGCGGGCAGCAGCATCGACGCTCATCTCCCACTGCGCCCCCGG-AGTGAGC 14056
QY 181 ACCGACCGCGGACGTGACGCCATGGAGGGCCACGTCCGCGGACGGATCACCCCTGGCT 240
DB 14055 ACCGACCGCGGACGTGACGCCATGGAGGGCCACGTCCGCGGACGGATCACCCCTGGCT 13996
QY 241 TCGGCCGAAGGTTCCGGTGTGTCGCCGCCAGATGCGGACGGCCTGGGGGCGCGCC 300
DB 13995 TCGGCCGAAGGTTCCGGTGTGTCGCCGCCAGATGCGGAAACGGCCTGGGGGCGCGCC 13936
QY 301 CGTCACTTCCGACAGGTCCGACAGCACTTGCGGCTTGCGCCCCCGCTGCGCTCGGC 360
DB 13935 CGTCACTTCCGACAGGTCCGACAGCACTTGCGGCTTGCGCCCCCGCTGCGGCTCGGC 13876
QY 361 GCTCTCCAGGAACAGGTCCGGACGGGCTCGGATATCTTCGGAGCATCTGTCGCGCGC 420
DB 13875 GCTCTCCAGGAACAGGTCCGGACGGGCTCGGATATCTTCGGAGCATCTGTCGCGCGC 13816
QY 421 CGCCTCCAGCCCCAGCTCTCGAAACGAGTGAACCGCCCCAGCACCTCGGAGAGCCGCGC 480
DB 13815 CGCCTCCAGCCCCAGCTCTCGAAACGAGTGAACCGCCCCAGCACCTCGGAGAGCCGCGC 13756
QY 481 CGTCTGCTGCTGGCGGTGATCGCTCGGGCCGACAGCGGTACGCGCGCTCCCTCGTG 540
DB 13755 CGTCTGCTGCTGGCGGTGATCGCTCGGGCCGACAGCGGTACGCGCGTCCCTCGTG 13696
QY 541 GCGGGCGCGGTCAGTGGCCCTGACCGCCATTTCGCGATGTGCGCGGATCGACAGGC 600
DB 13695 GCGGGCGCGGTCAGTGGCCCTGACCGCCATTTCGCGATGTGCGCGGATCGACAGGC 13636
QY 601 AACCGGGACGTGCGGTACAGCGCGGACACAGCCGCTCGACCCGGATGGCGGGCGCCA 660
DB 13635 AACCGGGACGTGCGGTACAGCGCGGACACAGCCGCTCGACCCGGATGGCGGGCGCCA 13576
QY 661 GGACACGTGTGGACATGAAGTCTGCGCGCAGGAGGTCCAGTCTAGCCCGGACTC 720
DB 13575 GGACACGTGTGGACATGAAGTCTGCGCGCAGGAGGTCCAGTCTAGCCCGGACTC 13516
QY 721 GCGTACGGCCCGCTCGTTCTCGCGCTGCGCGCGCTGATGAAGTCTGCGCGCCCGGTTTC 780
DB 13515 GCGTACGGCCCGCTCGTTCTCGCGCTGCGCGCGCTGATGAAGTCTGCGCGCCCGGTTTC 13456
QY 781 CCCCACCGGGACATGGACAGCTTACACAGGTGCGGACCGCGGCTCGCGGCGCGCGC 840
```

```
DB 13455 CCCACCGCGAGCATGGACAGCTTCCACAGGTGCGGACGCGCGCTCGCGCGCGCGCGC 13396
QY 841 CGCGAAACGCTGCTGCTGCTCGGCTCGGTGGGACATCTGTTGTCGACGAGGAACCGCGCGCGCAC 900
DB 13395 CGCGAAACGCTGCTGCTGCTCGGCTCGGTGGGACATCTGTTGTCGACGAGGAACCGCGCGCGCAC 13336
QY 901 CCCGTTGAGGGCCGGTCCAGGCCCGGGCGGCTCGGCGTACTCGGCCCGCGAGACTCGAC 960
DB 13335 CCCGTTGAGGGCCCGGTCCAGGCCCGGGCGGCTCGGCGTACTCGGCCCGCGAGACTCGAC 13276
QY 961 GTTTCGGGCGCGGTGACCGGTTCACCCGTTCCGGGCGCGCGGCGGAGGACTCTGACGGGACCGGT 1020
DB 13275 GTTTCGGGCGCGGTGACCGGTTCACCCGTTCCGGGCGCGCGGCGGAGGACTCTGACGGGACCGGT 13216
QY 1021 CCGGGCCAGACAGGTGGGCGACCTGACGGCCGACACACACCGGTTCAGCGCGGTTCACAGAAAT 1080
DB 13215 CCGGGCCAGACAGGTGGGCGACCTGACGGCCGACACACACCGGTTCAGCGCGGTTCACAGAAAT 13156
QY 1081 CACTCGGGGCTCCTCTCGGGCAGCGAGGCGGCGCTCCGAAACATACATATGAGGGGA 1140
DB 13155 CACTCGGGGCTCCTCTCGGGCAGCGAGGCGGCGCTCCGAAACATACATATGAGGGGA 13096
QY 1141 AGGCGACGATCT-GCCGCCGGGGCGCGAAACCGGCGATGTTGCGGCCCGGGGCGCGGTGCTT 1199
DB 13095 AGGCGACGATCTGCGGCCCGGGGCGCGAAACCGGCGATGTTGCGGCCCGGGGCGCGGTGCT-T 13037
QY 1200 CAGCCGAGAAACGCGGGCGCGGACAGCGTGTGTAGGAAGACGAGGCTGCGGCTCTGTATGC 1259
DB 13036 CAGCCGAGAAACGCGGGCGCGGACAGCGTGTGTAGGAAGACGAGGCTGCGGCTCTGTATGC 12977
QY 1260 CCGGTGACCCGCAAGGTGTCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGCTGGATC 1319
DB 12976 CCGGTGACCCGCAAGGTGTCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGCTGGATC 12917
QY 1320 CAGCACCGGCTGTGCAACTCCGCGTACCGGACGAAACGGGTCGCGCGGATCGAGGGCACG 1379
DB 12916 CAGCACCGGCTGTGCAACTCCGCGTACCGGACGAAACGGGTCGCGCGGATCGAGGGCACG 12857
QY 1380 AAGGGCGCGGACCGGTCACAGGACAGCGCGCTGCGGCTGCGGCTCGAGCAGCAGCATG 1439
DB 12856 AAGGGCGCGGACCGGTCACAGGACAGCGCGCTGCGGCTGCGGCTCGAGCAGCAGCATG 12797
QY 1440 CCGGTAAGTGTGTTGGGGCGCTGGGAAGAGGTCGCGTACCTGCTGCTCCACCGCAGT 1499
DB 12796 CCGGTAAGTGTGTTGGGGCGCTGGGAAGAGGTCGCGTACCTGCTGCTCCACCGCAGT 12737
QY 1500 CGCACAGTGTCTGTGTCGAACTCGCCGACAGGACCAAGTCTCTGTCGCGGTCGACCG 1559
DB 12736 CGCACAGTGTCTGTGTCGAACTCGCCGACAGGACCAAGTCTCTGTCGCGGTCGACCG 12677
QY 1560 GCGCGCGCGCGGACAGGGCGCGGTCGCGGCGACCGGATCGGTGGGAGTTCGCGAGTTCG 1619
DB 12676 GCGCGCGCGCGGACAGGGCGCGGTCGCGGCGACCGGATCGGTGGGAGTTCGCGAGTTCG 12617
QY 1620 CCGCGCATCCCGCGGTAGACTTGGAGGACTGCTGAAGCGCTGAGCGGTCGCGGAGTCCCGTGCA 1679
DB 12616 CCGCGCATCCCGCGGTAGACTTGGAGGACTGCTGAAGCGCTGAGCGGTCGCGGAGTCCCGTGCA 12557
QY 1680 GCGAGCGCTCGCGGCGCGGACAGCGCCCAAGTCTCACTGCTGCTGTAAGGGTTCGCGCGC 1739
DB 12556 GCGAGCGCTCGCGGCGCGGACAGCGCCCAAGTCTCACTGCTGCTGTAAGGGTTCGCGCGC 12497
QY 1740 CCGAACTTCAGCTGGGAAACAGGCCCATCTTCCACTTCAGCTCCGCGACCTCGCGCGACAG 1799
DB 12496 CCGAACTTCAGCTGGGAAACAGGCCCATCTTCCACTTCAGCTCCGCGACCTCGCGCGACAG 12437
QY 1800 CCGAGGTGCTGAGGTGCGAGGTGTAGTTCAGCGTGGCCATCAGGAAGTGTAGCCACCC 1859
DB 12436 CCGAGGTGCTGAGGTGCGAGGTGTAGTTCAGCGTGGCCATCAGGAAGTGTAGCCACCC 12377
QY 1860 GGCACGCGGTAGCGCGGTGGAAGACGAGCATCGCGGCTGACCGAGGCTTCGCGCGATC 1919
```

12376	GGCAGCCGCTAGCCGCGGTGGAGAACGAGCATGCCCGCTGACGACGGTCTCGCGCATC	12311
QY	AGCAGCGGATCGTGTCCGTGCCCGCTGGAGAAACGGGTGCTCGTGGGCGAG	1979
Db	AGCAGCGGATCGTGTCCGTGCCCGCTGGACCGGTGCGAAGAACGGGTGCTCGTGGGCGAG	12257
QY	ACGGCGGTGACCGAGAAACCGGTACACGCCCTTCGGTATCCAGCTGACCGGGAAACGGCTCC	2039
Db	ACGGCGGTGACCGAGAAACCGGTACACGCCCTTCGGTATCCAGCTGACCGGGAAACGGCTCC	12197
QY	TGCACCTCGTCCGGTGGACAAGCGCCATTCGGAACCGGCAATGCGGTTGTTCGATCGAG	2099
Db	TGCACCTCGTCCGGTGGACAAGCGCCATTCGGAACCGGCAATGCGGTTGTTCGATCGAG	12137
QY	TTGGCATTCGACGACGAAATTGATCAAACTACTGCTTCGGGCATGGGTCCCCCCCAGGAA	2159
Db	TTGGCATTCGACGACGAAATTGATCAAACTACTGCTTCGGGCATGGGTCCCCCCCAGGAA	12077
QY	TCATGTGATCCGAGCTGTTCTGTATGCGGAAACGTTAAGATACAGACTGAGCGGTTTTT	2219
Db	TCATGTGATCCGAGCTGTTCTGTATGCGGAAACGTTAAGATACAGACTGAGCGGTTTTT	12017
QY	TTTCTATCTTCCCGGGGGAGACATCAACAGAGGACGGCATGCCAAGCAGAACCGGG	2279
Db	TTTCTATCTTCCCGGGGGAGACATCAACAGAGGACGGCATGCCAAGCAGAACCGGG	11957
QY	CGATCCGACGCGGACGATCCTTCGACCGCGCGCGCAGGTCTTCAGAGAAGCAGGGCT	2339
Db	CGATCCGACGCGGACGATCCTTCGACCGCGCGCGCAGGTCTTCAGAGAAGCAGGGCT	11897
QY	ACCAAGTGCACGATCAACGAGATCCTCAAGGTGCGGGGTGACCAAGGGAGCCCTCT	2399
Db	ACCAAGTGCACGATCAACGAGATCCTCAAGGTGCGGGGTGACCAAGGGAGCCCTCT	11837
QY	ACTTCACATTCCTCCAGGAGAACTCGCGCTGGCGCTCTTCAGACCCACAGAACCCAC	2459
Db	ACTTCACATTCCTCCAGGAGAACTCGCGCTGGCGCTCTTCAGACCCACAGAACCCAC	11777
QY	CACAGGCGCTTCGGAGCAACCCCTCCGGCTGCAAGAACTCATCGACATGGGCATGTTGT	2519
Db	CACAGGCGCTTCGGAGCAACCCCTCCGGCTGCAAGAACTCATCGACATGGGCATGTTGT	11717
QY	TCGTGACACGCTTGCGCACGAAACGTGTGGCCCGGGCCGGCGCTCTCCATGGAAC	2579
Db	TCGTGACACGCTTGCGCACGAAACGTGTGGCCCGGGCCGGCGCTCTCCATGGAAC	11657
QY	AGCAGGCGCAGCTTCGATCGCCGAGAACCTTCGTCGCTGGCCACGAGACATCTCTGA	2639
Db	AGCAGGCGCAGCTTCGATCGCCGAGAACCTTCGTCGCTGGCCACGAGACATCTCTGA	11597
QY	AGCTGCTGAACACGAGGACGAGGAGTGTGTCGCCCATGTGCTCACCACCGACT	2699
Db	AGCTGCTGAACACGAGGACGAGGAGTGTGTCGCCCATGTGCTCACCACCGACT	11537
QY	CGGCCGATCTTACGTGGGCACGTTCCCGGGGATACAGTGTGTGTCACGAGCGTACGC	2759
Db	CGGCCGATCTTACGTGGGCACGTTCCCGGGGATACAGTGTGTGTCACGAGCGTACGC	11477
QY	ACTACAGGACCTCGAACACCGCTACGCGTGTGCAAGACATCTCTGCCGCATCG	2819
Db	ACTACAGGACCTCGAACACCGCTACGCGTGTGCAAGACATCTCTGCCGCATCG	11417
QY	CGGTTCCCTCGTGCTGGCCGCTCGATCTCTCCAGAGAGCGGAGCACGCTTCGGG	2879
Db	CGGTTCCCTCGTGCTGGCCGCTCGATCTCTCCAGAGAGCGGAGCACGCTTCGGG	11357
QY	CCGAACCTGGCACCGACCGGGAGGAGTGAACCGCGAGGCCCGCGCATCCGAGTACCGACC	2939
Db	CCGAACCTGGCACCGACCGGGAGGAGTGAACCGCGAGGCCCGCGCATCCGAGTACCGACC	11297
QY	GCGTGCACGAGCGGCGCAACCGGTGTCGCGTGTGTCGAGAAACATCTCTGCCGCATCG	2999
Db	GCGTGCACGAGCGGCGCAACCGGTGTCGCGTGTGTCGAGAAACATCTCTGCCGCATCG	11237

Qy	3000	CCCTCGGTACCGAGCGCTCGCGGTCAGAGATCGTTCGGCCTGTGGCATCGACGTACTG	3059
Db	11236	CCCTCGGTACCGAAGCGTGGCGGTCAGAGATCGTTCGGCCTGTGGCATCGACGTACTG	11177
Qy	3060	GCCGCGTATCCACCGGTGAGTCTCGAGGSCCAGAAAGCCACACGTCGGCGATGTGCTC	3119
Db	11176	GCCGCGTATCCACCGGTGAGTCTCGAGGSCCAGAAAGCCACACGTCGGCGATGTGCTC	11117
Qy	3120	GGGTCTGCCGATGCGGTGTAAACAACGAGTGTGCGGCCAGTCCGCGCGCGCTCGGGGT	3179
Db	11116	GGGTCTGCCGATGCGGTGTAAACAACGAGTGTGCGGCCAGTCCGCGCGCGCTCGGGGT	11057
Qy	3180	CTGCGCGCGCGTGCCTTCATGTCCGTCTCAGCAAAACCGCGGCCACCGCGTGTACCGT	3239
Db	11056	CTGCGCGCGCGTGCCTTCATGTCCGTCTCAGCAAAACCGCGGCCACCGCGTGTACCGT	10997
Qy	3240	GATCCCCCGTCCCCCAGTTCCTGCGCAGGCGAGCGTGTGTCACCGCACCCCTT	3299
Db	10996	GATCCCCCGTCCCCCAGTTCCTGCGCAGGCGAGCGTGTGTCACCGCACCCCTT	10937
Qy	3300	GGTCATCGGATATCCGATGGACTCGGCGAAACCGCGCGCGGTTCGGCGAGACGAGATGT	3359
Db	10936	GGTCATCGGATATCCGATGGACTCGGCGAAACCGCGCGCGGTTCGGCGAGACGAGATGT	10877
Qy	3360	GATGATCCGCGCGCGTCCGCGAGTCTGTTCACTCGTGTGAGACGAGAACAGCGGTGC	3419
Db	10876	GATGATCCGCGCGCGTCCGCGAGTCTGTTCACTCGTGTGAGACGAGAACAGCGGTGC	10817
Qy	3420	CCGGAAGTTGACGGCGACAGTCCGTGAGAGACTCTTCGTGTACTTCGGTGATCCCGTCC	3479
Db	10816	CCGGAAGTTGACGGCGACAGTCCGTGAGAGACTCTTCGTGTACTTCGGTGATCCCGTCC	10757
Qy	3480	CGAGCCGCTGAGCCCGCGTGTTCACACGAGATGTGAACTCGGCGCGCACTCCGAACTC	3539
Db	10756	CGAGCCGCTGAGCCCGCGTGTTCACACGAGATGTGAACTCGGCGCGCACTCCGAACTC	10697
Qy	3540	GCCCATCCCGCGTCAACCGCGCGTAGAGCGCGCGCTCACCCACGAGCCGAGTTC	3599
Db	10696	GCCCATCCCGCGTCAACCGCGCGTAGAGCGCGCGCTCACCCACGAGCCGAGTTC	10637
Qy	3600	GGCCGGATGGCAACGCTGTCCGCGCTGTCTCGGATGTCTTCGAGGTCCTCCGCGC	3659
Db	10636	GGCCGGATGGCCAAAGCCTGTCCGCGCTGTCTCCGGATGTCTTCGAGGTCCTCCGCGC	10577
Qy	3660	CGCGCTCGCTGCTGCGGTAGTGACTCCACAGAGCGCCCGTCCGGCGCGCGCAGCGAG	3719
Db	10576	CGCGCTCGCTGCTGCGGTAGTGACTCCACAGAGCGCCCGTCCGGCGCGCGCAGCGAG	10517
Qy	3720	GGCGATACCGGTCCGATGCCCGCTTCCCGCGTCCACAGGGCGGTCTTGCCTCCAG	3779
Db	10516	GGCGATACCGGTCCGATGCCCGCTTCCCGCGTCCACAGGGCGGTCTTGCCTCCAG	10457
Qy	3780	CGGTCTTCCATACCTCGTCCCATGTGTGACAGCATATCAGCCCCCGCGTGTGAGCGACC	3839
Db	10456	CGGTCTTCCATACCTCGTCCCATGTGTGACAGCATATCAGCCCCCGCGTGTGAGCGACC	10397
Qy	3840	CATGGCGGCGCTCGGCGGTTCGAATCGAGGTCACAGCTTACCTGTGACCGCGTCAGAC	3899
Db	10396	CATGGCGGCGCTCGGCGGTTCGAATCGAGGTCACAGCTTACCTGTGACCGCGTCAGAC	10337
Qy	3900	GGGCGCGAGTGCCTGGTTGACAGGCTCGGGCCAGATCGGCGCGCGGACCGGGAAACC	3959
Db	10336	GGGCGCGAGTGCCTGGTTGACAGGCTCGGGCCAGATCGGCGCGCGGACCGGGAAACC	10277
Qy	3960	GGCGCCGGTCAAGGGGTTCAGGGGTTCGCGGGGACCGGCCAGGCGCGGTACGGGACCGACGG	4019
Db	10276	GGCGCCGGTCAAGGGGTTCAGGGGTTCGCGGGGACCGGCCAGGCGCGGTACGGGACCGACGG	10217
Qy	4020	ATCGAGTCTGGGCGTGCACAGCGGGCACAGTCTTCGCGGGCCAGCTTCGACTCTGACGC	4079
Db	10216	ATCGAGTCTGGGCGTGCACAGCGGGCACAGTCTTCGCGGGCCAGCTTCGACTCTGACGC	10157

QY	4080	GTACACAGCCCGGTCCGCGCCAGTCTGAGCTGACGTGGCGCGCGGTGGGTGAGCG	4139
Db	10156	GTACACAGCCCGGTCCGCGCCAGTCTGAGCTGACGTGGCGCGCGGTGGGTGAGCG	10097
QY	4140	GTTGCGCCAGCGCGGAGGAGCGCGGAGGTCCGCGCGAGCAATCATGCGCGCGCGCGGTCC	4199
Db	10096	GTTGCGCCAGCGCGGAGGAGCGCGGAGGTCCGCGCGAGCAGCAGGCGCGCGGTCC	10037
QY	4200	GAACGCGCGCGCGGTCCAGGGTCTCTCCAGGACGTCTAGACCCGCCAACCCGCC	4259
Db	10036	GAACGCGCGCGCGGTCCAGGGTCTCTCCAGGACGTCTAGACCCGCCAACCCGCC	9977
QY	4260	CTGCGCGCAGCGCGGAGCGCGCCGCGCAGCTCCGCGGTGTCGCGTCCGCGCGCGAGGC	4319
Db	9976	CTGCGCGCAGCGCGGAGCGCGCCGCGCAGCTCCGCGGTGTCGCGTCCGCGCGCGAGGC	9917
QY	4320	GAGCGACCGGTAGCGCGCGGTACC	4346
Db	9916	GAGCGACCGGTAGCGCGCGGTACC	9890
RESULT 3			
AB011413/c			
LOCUS	AB011413	12070 bp	DNA linear BCT 07-AUG-1998
DEFINITION	Streptomyces griseus genes for Orf2, Orf3, Orf4, Orf5, AfsA, Orf8, partial and complete cds.		
ACCESSION	AB011413		
VERSION	AB011413.1		
KEYWORDS	Orf8; AfsA; Orf5; Orf4; Orf3; Orf2.		
SOURCE	Streptomyces griseus		
ORGANISM	Streptomyces griseus		
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyceinae; Streptomycetaceae; Streptomyces.		
AUTHORS	1 (sites)		
TITLE	Umeyama,T.		
JOURNAL	Open reading frame encoded around afsA gene		
REFERENCE	2 (bases 1 to 12070)		
AUTHORS	Umeyama,T.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-FEB-1998) Takashi Umeyama, University of Tokyo, Department of Agriculture and Life Sciences, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan		
FEATURES	(E-mail:aa67103@nongo.ecc.u-tokyo.ac.jp, Tel:++81-3-3812-2111) Location/Qualifiers		
source	1..12070		
	/organism="Streptomyces griseus"		
	/mol_type="genomic DNA"		
gene	/db_xref="taxon:1911"		
	complement(2660..3511)		
CDS	/gene="orf2"		
	complement(2660..3511)		
	/gene="orf2"		
	/function="sensor histidine kinase"		
	/codon_start=1		
	/transl_table=11		
	/product="Orf2"		
	/protein_id="BAA32130.1"		
	/db_xref="GI:3401947"		
	/translation="VAAVPAPHPPEAAPAAARLTAAYTLITVUGLACLSMLVIR		
	TDDRMDAEDVMERRASVASLYYEDDIRSAGPAEHSLSVLIVIGAGALTALA		
	ALTGHVLSRSGPAWEALQEORLULADAEHLRTPVAVMRGSGVEAAGAGGLGQL		
	PRIIRAADMAVDVNLTRGLEAGTSGYRSRLRLDOLVEEVCAELPEGGPGLES		
	LEESVEADAALVRVAVRNLLDNAVRHGPSRGAAGRTAGDRVTRGVRAADRPAY		
	APARLPS"		
	complement(3523..4197)		
CDS	/function="two-component regulator protein"		
	/codon_start=1		
	/transl_table=11		
	/product="Orf3"		
	/protein_id="BAA32131.1"		
	/db_xref="GI:3401948"		
	/translation="MRVLLVEDDDDLRDVIGNGLRDGGFAWDCASDWPEADVILLHLSA		

YDCVILDRMVPSGDTLPLEGERRRAGMSPVPLCLTALDSILDERLRLGLSGSADDYDLAK	
FSMRLLVLRGSLRSGARSLFSFGCCADVMVDAHRRVGRGGLVLSLSPKEVAVYLQ	
LLVHRDITVTRTGLLEHCWDEADPVSNVVDVAVAGLRKRLGSPGLVHTVTRGGQFLLS	
ADPGPS"	
CDS	complement(4333..5250)
	/function="methyltransferase"
	/codon_start=1
	/transl_table=11
	/product="Orf4"
	/protein_id="BAA32132.1"
	/db_xref="GI:3401949"
	/translation="MATQALGTALRLGVDFRICTGELRAEALAGSLGTHPRATLRLR
	ALAGLQLSEPPGAFRTTAAGNVLRADAPGTVMARMFTDPVMLRGWLLDESVRT
	GETTFDVFQDFEHLRHEPESAAFNAMSQGLRLTAETVPHHYDFGFRQRLVDIG
	GGDTFTLAIIRAHQEPGRVLPDTAEGLAQAPRELAREGLDGRVTLTGDFEASAPAG
	GDLYLNSIIHDWDDAVRHLRHIRDVIIPHGSLLIIVEPVLPAIVPADRPDNYLSDL
	NMLVNVGGRERTADDFAALCTAGGFACGA"
CDS	complement(5450..6361)
	/function="regulator protein"
	/codon_start=1
	/transl_table=11
	/product="Orf5"
	/protein_id="BAA32133.1"
	/db_xref="GI:3401950"
	/translation="MSELPLDQVRTLLAVVDEGTFDSAATLRLTPSAVSQRVKALEQ
	RTGRVLLMRTPKVRPTESGEVVALARLARLEHQAALGMSGPGEPTLLPAINVSDS
	LATSWQPCACRSWDSATTPGPGPYGRPAGGAGDGRDLVAGGDLGFGPAAG
	RMRYLPVSPGFADRWLRDRDGTALRELIGEAPVVCDFDRDDLQDAFVRLGPGARPS
	ARRHLVPTSGEFANAVASGMGWGPVEVQAEPLLSDGLRVRLAPEPTVDVPLYWQQWK
	LESPALAAVAEAAEALDEAPPN"
CDS	7486..8391
	/codon_start=1
	/transl_table=11
	/product="AfsA"
	/protein_id="BAA32134.1"
	/db_xref="GI:3401951"
	/translation="MDAAEVVHPVGIEMVHRTPEDAFPNNVRLGRDFAVEAVLP
	HDHPPAPVGDLDHPLVABAMQAAALAPHAGYGPLGVHFLLTLDYVCHPEHLG
	VGEPTEIGLEVCSDLKWRAGLPAQGRVGMVAGHGRDLAATGAATVAATFSTPKAVRRM
	RGDVVEIGSLPTAPVPASGRARVEDVVLSTGREGWELRVDRTHPTLFFORPND
	HVRMMLLEAARQAACLVAGPAGIVPVEARTFRHRYSEFGSPCMGAVVQGADEDTV
	TPTVGHQDGETVFSTVLSGPRAGH"
CDS	10558..11598
	/function="alcohol dehydrogenase"
	/codon_start=1
	/transl_table=11
	/product="Orf8"
	/protein_id="BAA32135.1"
	/db_xref="GI:3401952"
	/translation="WTTVAAYAPRAKAPLERTTIERRPVGEFDILIDIKFAGICHSD
	IHQARDWGEGIFPMVFGHEIAGIVAGTGPVTRFKVGDVGVGMDVSCGTCDACLM
	GREQHCAGNTQTNALDRNGEPTVGGYSTHLVTEKFALSIPEGIALDEAAPLLCAG
	ITTSPLRRMGAGFKKVAVVGLGLGHMAVKIAHALGAETVLSQSLRKXDDGLKLG
	ADHYATDPTPTPEELAGTDFVILSTVAPLDYFAGLALLTEGLTVNVGAPPEPVSIL
	NLFSLHGNRSIAGSAGTIBETQEMLDFFCAVHGLGAEIEVIGAGQVNEAVERVLASD
	VRYRPFVIDTATI"
ORIGIN	
Query Match 20.7%; Score 901.2; DB 1; Length 12070;	
Best Local Similarity 68.6%; Pred. No. 6e-69;	
Matches 1361; Conservative 0; Mismatches 508; Indels 115; Gaps 4;	
Qy	246 CGAAGCTTCGCGTGTCCGCCGACAGATCGGAAACGGCTGGCGGCGCGCCGCTCA 305
Db	9238 CGAACCGCTTCGCGTGTCTCGGCCGACACAGTCCGCGAACGGCTCGCGGACGCCCGCTCA 9179
Qy	306 CTTCCCGCAGCTCGGCACGACCTTGGCGCTTCGGCCCCCGCGTCCGCTTCGGCGCTCT 365
Db	9178 GCTCTTGACCGTGGGGCAACACGACGACGCTTCGCTCCGCTCGCGTCCGCGCTCT 9119
Qy	366 CCAGGAACCGCTCGCGACGGGCTTCGGATACTTCCGAGCATCTGTCGCGCGCGCGCT 425
Db	9118 GCAGGAAGGCTTCGCGACGGCGGTGGGTACTTTCGCCATCAGGGCCGTTCCGCGGCGGT 9055


```
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein D"
/protein_id="BAC76464.1"
/db_xref="GI:30698351"
/translation="MTGLPEPTVRLRQVRLTRDPARGELALLPERVVVLDNTAAAV
LAHCDGTTSLAGIVRLAEVGVSAEDVRELLLLAQRVVLDLG"
/complement(7171..7899)
/notes="ORF7 (242 aa), lankacidin biosynthesis protein
similar to AE004625-5 Pseudomonas aeruginosa
pyrroloquinoline quinone biosynthesis protein C (250 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein C"
/protein_id="BAC76465.1"
/db_xref="GI:30698352"
/translation="MSMSTVREVAAPWSEAEPRORLHALESSYMDRHPFRRMHGELL
DEGELRLWAANWYORCLPKQDAIVANCPLEVRROWLSRIYVLDHADACAGAEK
WURLABVGLRDEVDHRLVLAGTRFAVDYDFARRPWLAAASGLTFLSPGLL
AHLRLREHYPIWIAEFGFEYFARIEVVGPGRSLDLDVAHVAVRSQREACVRALA
FKCRVLAVLDSLDYHTNGATRS"
/complement(7926..8030)
/notes="ORF8 (34 aa), lankacidin biosynthesis protein
similar to AL603642-197 Sinorhizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein A (31 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein A"
/protein_id="BAC76466.1"
/db_xref="GI:30698353"
/translation="MRTSGKELPAKKAMHRPFDVITDGMETVAFPSR"
/8420..10057
/notes="ORF9 (545 aa), possible lankacidin resistance
protein
similar to AL133424-19 Streptomyces coelicolor putative
ABC transporter ATP-binding protein (544 aa)"
/codon_start=1
```

Query Match

Best Local Similarity 62.8%; Pred. No. 1,8e-40;

Matches 981; Conservative 0; Mismatches 535; Indels 47; Gaps 5;

```
QY 2260 CATGGCAAGCAGGACCGGGCGATCCGACGGGCGAGATCCTGGAGCGCGCGGCGCA 2319
DB 143125 CGTGGCACAGCAGGAACGGGCGATCCGACCGCGCGCGGCTTCTCGAAGCGCGACCCAC 143066
QY 2320 GGTCTTCGAGACGAGGGCTACCAAGTCGACGATCAGGAGATCTCAAGTGGCGCG 2379
DB 143065 GGTATTGCGCGAGCAGGGTAGCGCCGCGACCGTGGCGGACATTTCTCAAGTGGCGCG 143006
QY 2380 GGTGACCAAGGAGCGCTTACTTCCACTTCCAGTCCAAAGGAAGAACTGGCGTGGGCGT 2439
DB 143005 CTTGACCAAGGAGCGCTTACTTCCACTTCCAGTCCAAAGGAAGCGCTTGGCGGCGCAT 142946
QY 2440 CTTTCGACCGCCAGGAACCAACAGCGCGTTCGAGAGCAACCCCTCCGCGTGCAGAACT 2499
DB 142945 CTTGGAGGCGCCAGGTGCGCAGCAGCTCGTTCGCGACGCA---GCTCAAGCTCCAGGAATG 142889
QY 2500 CATCGACATGGGCATGTTGTTCTGTTCACCGTTCGCGCAAGACGTGCTGCGCGCGGCGCG 2559
DB 142888 GGTGGACGCGGGCGATGACCTGGCGCACCGAGCTGCGCGGGGACCCCGTGTGCGCGCGG 142829
QY 2560 CGTGGCGCTCTCCATGACACAGCGCGACCGTCTCGATCGCGCGAGGACCCCTTCGCTCG 2619
DB 142828 CGCGCGCTCTTCGGCGAGGACACACCGGAGCGAGC---AGCAGGAGCGCGCTTCCCGAC 142772
QY 2620 CTGGCACAGACACATCTCTGAAGCTGTGAACAGGCGCAAGGAGAAACGGTGAAGTGTGCC 2679
DB 142771 CTGGATCGCTTCTCGCGCTCGCTCTCTTGAGCAGGCGGAACGGGAGGTCTCTGGG 142712
QY 2680 CCATGTGGTCAACACCGACCTCGGCGCATCTCTACGTGGGACGTTTCGCGGGATACAGT 2739
DB 142711 CCACATGACGCGCGCGAGACCGCGCATGCGTCTCGGCTCTCTCCACGCGCATCAGCT 142652
```

```
QY 2740 CGTGTCCAGACGGTACGGACTACAGACCTCGAAACACCGCTACGCGCTGCTGCAGAA 2799
DB 142651 CTTCTCCAGTTGCAGACGAATCGGCGGACATCAGCAGCGTGTCTCGCGCTGTTCGG 142592
QY 2800 GCACATCTGCGCGCATTCGGGTTCCCTCCGTGCTGCGCGCTCGATCTCTCCGAGGA 2859
DB 142591 GCACGTCTCTCGGCGCGTGGCGGTCCCTCGGTGCTGTGTGGCGGTGGACACCGCCCCGA 142532
QY 2860 GCGCGAGACCGCTTCGCGGCGGAATCGGCACCGCG----- 2897
DB 142531 CCGGCGGCGCGCGTCTGTCGCGAGCTGGAGGCGATGGCGGCGCAGCCGCGGCTGGC 142472
QY 2898 -----GGAAGACTGACCGCGGAAGCGCCCGACCGCATACC--GACCGCGC 2942
DB 142471 CTCGCTGTGAGCTGAACCGCGCGCGCGCGCGCGCGCGACCGGGGACCGGTCTATGG 142412
QY 2943 GTGCCCGAGCGCGCGACCGCGGCGCGCTTACGCGGCGCGCGCGCGCGCGCTAGTCTGCCC 3002
DB 142411 CCGCGCGCGCGCGGATCAGGGCGCGCGATCAGTCTTTTGTCCAGGCTCTTTGCCGAGGC 142352
QY 3003 TCGGTACCGAAGCGTGGCGGCT----CAGAGAAATCGTTCCGCTCTGTGGCATCGAGTACT 3058
DB 142351 CTTGCCCAACCATCTTTTGGGAGTCTTACAGCGCGCGCGCGCGCGCGCTCCAGGCACT 142292
QY 3059 GCGCGGTGATCCACCGTGAAGTCTGGAGGCGCAGAAAGCCACACGCTCGGCGATGTCTGT 3118
DB 142291 GACCGGTGATCCACCGGAGTCTCCGAGGCGAGAAAGCGCGATGTCTGGCGATGTCTCG 142232
QY 3119 CGGGTCTCGGATCGGTTGAACACGGAGTTGGCGCGCGAGTCCCGCGCGCGCTCGGGGG 3178
DB 142231 CGGTCTCTCCATGCGCGGAGAGACCGAGGCGCGGTGCGCGCGCTCGCGCTCGGGTG 142172
QY 3179 TCTGCGCGCGCGTGTCTATGCTCTTCCAGAAACCGCGCGCGCGCGCGCGATGTGACCG 3238
DB 142171 TGGCCACAGCTCGCGGTTCAATGCTTCAAGAAATCCGCGCGCGCGCGCGCTTCCCG 142112
QY 3239 TGATCCCGCTTCCCGCGTTCCTGCGCGCGCGCGCGCGCGTGTCCACCGCGCGCT 3298
DB 142111 TGATCCCGCGCGCGCGCGCTTGGCGCGCGCGCGCGCGTGTCCAGCGCGCGCT 142052
QY 3299 TGGTCACTCGGTATCCGATGAGCTCGGGGAACGCGCGCGCGCGCGCGCGAGATGT 3358
DB 142051 TCGCCATGAGTGAAGGATGAGCTCGGNAAGCGCGCGCGCGCGCGCGCGCGCGCGTGT 141992
QY 3359 TGATGATCCCGCGCGTTCGCGCGTTCAGTTCGCTGCTGGACCGAGAACAGCGGTG 3418
DB 141991 TGAAGCATCGTCCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 141932
QY 3419 CCGGAGCTTGAAGGCGACAGTCCGTGGAAGACCTCTCGGTGACTTCCGTGATCCGTC 3478
DB 141931 CCTTGGCGTGAAGGCGCGATCAGTCTCGAAGTCCGCGCGCGCGCGCGCGCGCGCG 141872
QY 3479 CCGAGCGCTGACCGCGCGTGTGTTTCCAGGATGTCCGAATCTCGGCGCGCGCGCGCG 3538
DB 141871 CCGGTGTGTGTTCCCGGTGTTTGAACAGAGATGTCCAGGCGCGCGTTCGCGCGCGTGT 141812
QY 3539 CCGCCATCCCGCGCTGCGAAGCGCGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGT 3598
DB 141811 CCTCAGCGCGCGTTCGATCGTCTGTAAGTTCGCGCGCGCGCGCGCGCGCGCGCGCG 141752
QY 3599 CCGCGCGAGTGGCAGACCGCTGTGCGCGCGTGTCCGATGCTCTCGAGTGTCTCTCGCG 3658
DB 141751 GGGCGTGCACCATGACGCGCGCGACCAACCGCTGTCTCGGTGATCTCGCGAGCGTGTCTCG 141692
QY 3659 CCGCGCGCTGCTGTGCGGTGAGTGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3718
DB 141691 CCGCGGTCTGTGCGCGCGGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 141632
QY 3719 GGGCGATACCGGTTCGATGCGCGCGCGCTTCCCGCGGTCCACGAGGCGCGTCTTGGCCCTCCA 3778
DB 141631 GGGCGATGCGCGCGCGCTATGCGCGCGTCCCGCGCGTGTGACGAGGCGCGTCTTGGCCCTCCA 141572
QY 3779 GCG 3781
```


Db 2435 GAGCTGGTGATGGCGCCGAGCGCTCGCGCGCGCTGTTGTTACGAGGATGTCGAACTC 2376
QY 3522 GGGCGGCACTCGAACTCGCCATCCCGCGGTGCAACCCCGGTAGAGCGCGCGCGTC 3581
Db 2375 CCGCGCCGCGCCGCTCCGGAGGCGCGCTCGAGGCGCGGTAGAAAGTGTGACGCGC 2316
QY 3582 ACCACGACGCGAGTTCGGCGCCGATGCCAACCTGTCCGCGCTGCTCCGATGGT 3641
Db 2315 GTCCGGCGCCGAGTTCGGCGCGAACGCAAGCGCGCGCCGCCCCCTCGATGAT 2256
QY 3642 CTCGACGCTCTCTCGCGCGCGCCCTCGCTGTGTCGTAGTGAATGCCACGAGCGCCCC 3701
Db 2255 CTCACCGTCTCCAGGCGCGCTTCTGTTGTCGCTAGTGAAGCGGACACAGGGCGCC 2196
QY 3702 GTCCGCGCGACGCGACGGCGGATACCGGTCCGATGCCCGGCTTCCCGGTCACCG 3761
Db 2195 GTCCGTGGCAAGTCTCGCGGATGCCCTCTTCGATGCCGGGCTCGCCCTGTGACGAT 2136
QY 3762 GCGGTCTTGCCTCCAGG-----GCTTTCATACCTCGTCCCATGTGCACGCATAT 3814
Db 2135 GCGGTCTCTCGCGGCAGCGGACATGTGCGTGTGCTTACCTTCTGTTACGGAATGG 2076
QY 3815 CAGCCCCCGGTGAGCGACCATGCGGCGCGCTCGCGCTTCGAATCGACGGTCA 3874
Db 2075 TCGAGGAGTCCGTGATCAACGCGCGTGTTCGCGAACCCGTTAAGCCGCGCCCG 2016
QY 3875 CAGCCTACTGTGACCGGTGACAGGGCGGAGTGCGCGGTGTTGACGGCTGGGGCCA 3934
Db 2015 GGTTCGACGCGAGTCCCGCCAGGCGCGTGTCTCGCGAACCAGTGTGCGTGGCGAGCA 1956
QY 3935 GATCGGCGCGCGCACGGGNAACCGCGCGGTGAGGGGTGAGGGTCCGCGGACCGC 3994
Db 1955 TGTCCAGTGTGTCCTCCGCAAGGGAGGGCGCGGAGCAGCTGCGGGCGGTGCGGAGGT 1896
QY 3995 CCA--GGCGGTGAGGCAACGACCGGATCGAGGTGCGGCGGTGCGACGCGCCACCA-GT 4051
Db 1895 CCACGCGCGTCCCGGGGAGGCCGTGCGGTTGACGAGGACTCGCGGCGCGCGAGGA 1836
QY 4052 CCTCGCGCGCACGCTCCGACTC-GTACGGGTACGAGCGCGGTGCGGCGCGAGTCTGAGC 4110
Db 1835 GGTCCGCGACCGGCAACCGGTTCCGGATGTTGACGTAGTACGTTCCCGGCTCCGTCGCGG 1776
QY 4111 TGGACGTGCGCGCGGTGTTGAGCGGTTGCGCCAGGCGCGAAGCGGGAGGTGCG 4170
Db 1775 GGGAGGCGCGCGCGGAGCGCGCGCCCGCAGGCTGTGACGTCGATCAGGGAGTGCA 1716
QY 4171 GCGGCGAGCATCATGGGCGCGCGGTGCAAAACGCGCGCGCGCGG 4217
Db 1715 GCGAGGACCGCGCTGAGCGGTCCCGGACGCGCGGAGGAGCGCG 1669

RESULT 7
BA000030_43/c
WPCOMMENT

Sequence split into 91 fragments LOCUS BA000030 Accession BA000030

Fragment Name	Begin	End	Accession
BA000030_00	1	110000	BA000030_00
BA000030_01	100001	210000	BA000030_01
BA000030_02	200001	310000	BA000030_02
BA000030_03	300001	410000	BA000030_03
BA000030_04	400001	510000	BA000030_04
BA000030_05	500001	610000	BA000030_05
BA000030_06	600001	710000	BA000030_06
BA000030_07	700001	810000	BA000030_07
BA000030_08	800001	910000	BA000030_08
BA000030_09	900001	1010000	BA000030_09
BA000030_10	1000001	1110000	BA000030_10
BA000030_11	1100001	1210000	BA000030_11
BA000030_12	1200001	1310000	BA000030_12
BA000030_13	1300001	1410000	BA000030_13
BA000030_14	1400001	1510000	BA000030_14
BA000030_15	1500001	1610000	BA000030_15
BA000030_16	1600001	1710000	BA000030_16

1700001 1810000
1800001 1910000
1900001 2010000
2000001 2110000
2100001 2210000
2200001 2310000
2300001 2410000
2400001 2510000
2500001 2610000
2600001 2710000
2700001 2810000
2800001 2910000
2900001 3010000
3000001 3110000
3100001 3210000
3200001 3310000
3300001 3410000
3400001 3510000
3500001 3610000
3600001 3710000
3700001 3810000
3800001 3910000
3900001 4010000
4000001 4110000
4100001 4210000
4200001 4310000
4300001 4410000
4400001 4510000
4500001 4610000
4600001 4710000
4700001 4810000
4800001 4910000
4900001 5010000
5000001 5110000
5100001 5210000
5200001 5310000
5300001 5410000
5400001 5510000
5500001 5610000
5600001 5710000
5700001 5810000
5800001 5910000
5900001 6010000
6000001 6110000
6100001 6210000
6200001 6310000
6300001 6410000
6400001 6510000
6500001 6610000
6600001 6710000
6700001 6810000
6800001 6910000
6900001 7010000
7000001 7110000
7100001 7210000
7200001 7310000
7300001 7410000
7400001 7510000
7500001 7610000
7600001 7710000
7700001 7810000
7800001 7910000
7900001 8010000
8000001 8110000
8100001 8210000
8200001 8310000
8300001 8410000
8400001 8510000
8500001 8610000
8600001 8710000
8700001 8810000
8800001 8910000
8900001 9010000
9010000 9110000

BA000030_90 9000001 9025608									
Continuation (44 of 91) of BA000030 from base 4300001 (BA000030 Streptomyces avermitilis)									
Query Match		8.7%	Score 377.2;	DB 1;	Length 110000;				
Best Local Similarity		64.1%;	Pred. No. 1.8e-24;						
Matches 568;		Conservative 0;	Mismatches 318;	Indels 0;	Gaps 0;				
Qy	227	GATCACCCCTGGCTTCGCGCGAGGCTTCGCGTGGTCCGCGCCCGCCAGATGCGGACCGC	286						
Db	62443	GAGCATGCCCGCATTCAGCGGAAACGGTTCCTGTGTCGCTCGCGCCACTGTGCGTAGGAG	62384						
Qy	287	CTGGCGGCGCGCGCCGCTCATCTTCGCGACGCTCGGCACGACTCGCGCTTGGCGCCCGCC	346						
Db	62383	CGTGTGGGTGGCCCGTGAACCGGGCGAAACGTGGGGTCGGTCTGTCGCTTTGGCGCCGAGC	62324						
Qy	347	CGCTGCCGCTCGGGCGCTCTCCAGAAACGCTTCGCGGACGCGGCGCTTCGGATACTTCCGAGC	406						
Db	62323	AGCTGGCGTTGTGCGCTCAGAGACGAGTGGCTGCACCACTTGGGCGGGTAGCGCGCGCC	62264						
Qy	407	ATCTGCTCGCGCGCGCTTCACGCCCGCTCTCGAAACGAGTGACCGCCCGCCAGCAC	466						
Db	62263	CACCGGGCGGAGCTCTGCTGTGTGTGTCAGTTCGCGGACAGCAGCGGCGGCGGAGGACC	62204						
Qy	467	TCGGAGAGCGCGCGCTCTGCTGCTGGCGGTGATCGCTTCGGCGCCCGGACAGCGGTAC	526						
Db	62203	GCGCGCAGGTGCGCGCTCTGTTGCTTCGCGGAGACCGCTTCGGGCGCGGTGAGGGGATAC	62144						
Qy	527	GCCGCTGCTCTGTGGCGCGGCGGGTCAGTGCCTTGACCGCCACTTTCGCGGATGTCGCGC	586						
Db	62143	ACCGGCGCTCTGTGGCGCGGTCGAGGTTCAGCGCGGTTCACCGCGACTTCGCGGATGTCACGG	62084						
Qy	587	GGATCGAGCGAGGCAACCGGGACGTGCGGTACAGCGCGGCGACCAACCGCTCGGACCGG	646						
Db	62083	GGGTTCGACGAGCGGCTTCTCCGAGCGGGGTACAGGCTCGCACGTGCTTCCCGCGG	62024						
Qy	647	ATGGCGGGCGCCAGACAGCGTGTTCGACATGAGGTCTTCGCGCGCGCGGAGGCTCCAG	706						
Db	62023	ATGGAGTTCGCGCCAGGACAGGGTGTTCGACATGAGGCGCGCGGTTCGACGAGGCTCCAG	61964						
Qy	707	TCTAGCCGCGACTCGCGTACGGCCGCTGCTTCTCGCGCTTCGCGCGCGGTGATGAAGTCG	766						
Db	61963	TCGAGCCCGGAGGACGCGAGGCTCTCTTCGCACTTCGCTGCCACCGGTGATCAGTCTG	61904						
Qy	767	TCGCGCGCCGCTTCCCGCACCGCGAGCATGACAGCTTACAGGTGCGCGGACCGCGGCC	826						
Db	61903	TCGGCGTCTGCTCACCGCGCGGTACGAGGACAGCTTCACCATGCGGTACGTACGCGCC	61844						
Qy	827	TCGCGCGCGCGCGCGGAAACGCTGCTGCTCGGCTCGGTGGCGACTGTTCTGTGACGAGG	886						
Db	61843	GCCACCGCGCGCGCGGAAACCGCGCTCTCTGACCGCGCGGTTCGACGCTGACCGAGG	61784						
Qy	887	AACCGCGCGCGCACCGCTTGAAGGCGCGGTTCAGAGCGCGGGCGGTTCGCGGTACTTCGCC	946						
Db	61783	AACCGCGCTTCAGCGCGCTTTCAGTGC CGCGAGCAGCGAGCGGTGCTCTCTGTACGCTCG	61724						
Qy	947	GCGCAGACTTCGACGTTTCGGCGCGGTGACGCTTACCGCTTCGCGCGCGCGCGGAGACT	1006						
Db	61723	GTGGCGACTTCGACCGCGGGCCCTTCGACCGGTGAGCGCGCGGTTCGCGGACCAAGG	61664						
Qy	1007	CTGACGGGACCGGTTCGGGCGCAGAGGTGGCGGACTTGACGGCGCGACACACCGGTACG	1066						
Db	61663	CGCACCGCGCGGACCGCGCGAGTGTGTGTCACCTTCGCGCGCGCCACACGCGCGGTGCG	61604						
Qy	1067	CCGCTCACAAGAACTCACTCGGGGCTCTCTTCGCGGAGCGAGGACG	1112						
Db	61603	CCGCTCACAAGAACTCACTCGGAGCGGCTCTCTTCGCGCGCGCGCGG	61558						

RESULT 8
LOCUS AY228176S2 5495 bp DNA linear BCT 02-APR-2003
DEFINITION Streptomyces WP 4669 PD 116740 angucycline type II polyketide synthase gene cluster, partial sequence.

ACCESSION	AY228177	GI:29469247
VERSION	AY228177.1	
KEYWORDS	2 of 2	
SEGMENT	Streptomyces sp. WP 4669	
SOURCE	Streptomyces sp. WP 4669	
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	
REFERENCE	1. (Bases 1 to 5495)	
AUTHORS	Yin, X.H., Yoon, E., Mahadevan, B. and Proteau, P.J.	
TITLE	Molecular cloning and sequence of the PD 116740 angucycline type II polyketide synthase gene cluster from Streptomyces WP 4669	
JOURNAL	Unpublished	
REFERENCE	2. (Bases 1 to 5495)	
AUTHORS	Yin, X.H. and Proteau, P.J.	
TITLE	Direct Submission	
JOURNAL	Submitted (31-JAN-2003) Department of Pharmaceutical Sciences, College of Pharmacy, Oregon State University, Pharmacy Building Room 203, Corvallis, OR 97331-3507, USA	
FEATURES	Location/Qualifiers	
source	1. .5495	
	/organism="Streptomyces sp. WP 4669"	
	/mol_type="genomic DNA"	
	/strain="WP 4669"	
	/db_xref="taxon:224730"	
	/note="adjacent to AY228176; approximately 50 bp gap between adjacent sequences due to extremely high GC content"	
gene	50. .835	
CDS	50. .835	
	/gene="pd2E"	
	/gene="pd2B"	
	/note="pd2E"	
	/codon_start=1	
	/transl_table=11	
	/product="putative ketoreductase"	
	/protein_id="AA065365.1"	
	/db_xref="GI:29469255"	
	/translation="MSDKHERVALVSGATSGIGLAAARLLAQGHVFLGSRDADNVA RTVXELTABGLDVGDTLDVRSALSIESFVQAAVDRFGPDVLDVNNAGSGGGVTADI DELWYDVETNLNSVFLRTAVLTTCGMRHKNRRIINIASTAGQGVVLGAPYSAS KHGVGFTKALGNELAPGTITVNAVCPGVETPMAQVRVAGYAAAYDTTDAIMEKFT AKIPLGRVSTPEEVAGLVGLASDTAAASITAAQLNVCGLGNF"	
	877. .1821	
	/gene="pd2D"	
	877. .1821	
	/gene="pd2D"	
	/note="pd2D"	
	/codon_start=1	
	/transl_table=11	
	/product="cycloase-like protein"	
	/protein_id="AA065366.1"	
	/db_xref="GI:29469256"	
	/translation="MSASELREVEHEITIGAPADVRLIADVANNPRIFPTIYVDH LERGSDELIRIATANGEPKNWTSRRVLDREALRITFRQVSTTPVATMGSTWLEP QGDGASRVLLHDYTDVDDPAGLKWIDVARNRSSELASLTKVLELAASABITFS FEDTVYDGSADYDFVNEAGLSERLPHVATVRLTSDPSGLQLEMDTRADGSGTH TTSRVCLQCKIAYKQITLIPALMLHTGYWTFDTDTGTSTASSQHTVVLRENIEK ILPGGAVAEAREYVRSALSTNSRATLGHAKAYAEARR"	
	1824. .4118	
	/gene="pd2OR"	
	1824. .4118	
	/gene="pd2OR"	
	/note="pd2OR"	
	/codon_start=1	
	/transl_table=11	
	/product="oxygenase reductase-like protein"	
	/protein_id="AA065367.1"	
	/db_xref="GI:29469257"	
	/translation="MAGADAATVDVVVVGAGPVLMLAGELRGADVTVLRBLDPTT TESRSTLHARTMELLKORGLDGDGDFDPLPHEPRGHFGGLPMDLTLPSYPQGW KYPQVTRSELLGARAARLGVRLRGHETGLTQDQGVTSARHAGRPVRFTRARYVG CDGNSAVRRRLAGVDFPGDDAERELIRADVGIDIPNRRFQRLPVGLAIARAAGDGT	

RVVHRFTGPPRDEARFADIVEAWREVTGEDIAGHTPLWVNSFGDASROARVVRGR VLLAGDAHQOMP1GGQALNLGLQDAVNLGKLAATVRGRAPEGLLDTYHDERHACGR RVLSNIRQAARLLGGPEVEALRSVIGBELTSYERVRAHLAGTISGLDVRHGTGCGPTL AGTLRPAALGATLGGPSTDLREARGLLLVLTDAPEPPARVARPWSGAVDTVAARP AEAGLPEAAALLVRPDGYLAWAGDTTEGLDAALARWFTPGPTAPRHTPHRQHVVR RTKMNLGTFTALVTYSGRGMGTAORLARDGALVAHYGTNEDDAAEVAVAGIERDG GRAFAARLGAQDVHLEFLALROGLKERTGSTELNLLVNNACVWGVSVPKEDTTPDM FDRI FAVNARAPPFLVQBALTNPDGGRI VNISSGLTRPANPDEIAVAMSGAVBOLA LHYKALGPRGITVNSVAPGITRNGNLFDIPEAVEAUAALUSTFNRVGEPODVADVVA FLVSDDARWVTGFSFVDSGGTLLG"		gene	4123...>5495
/gene="pd2j"		CDS	4123...>5495
/gene="pd2j"			
/note="pd2j"			
/codon_start=1			
/transl_table=11			
/product="putative transporter"			
/protein_id="AA065368.1"			
/db_xref="GI:29469258"			
/translation="MTRTHPAPVGAADGRWSRLWGLLLVLAGNMLIDALEVSVAVV ALPSIGDGLGPTLDVHWMTGFVAGFGGLLFGARVVSLLGRRRVLAALLAFAAS ALSALAEPLLVATRLVKGFCAALTAFTGLAI TATAPQGPARDRAVSVVAFGAG FTAGLLSGMLVTEVSWRLAFAPVAVLVLPALALRLIPDPPAPARTSPRRYDLTGA ATFATAILLVYVIGAVPAHGWTSFRAAGPLTAAVLLAVFAVVERTTSAPLVOPAA RPPLIRSLGAACLSVYLGLLVLTVMOTVAGVBPSTALALLPAGPPALTAAP GLVARFAPRLIAGAACAFAGLQLSPAWPLHYPARLLPTLLLVGAAPVLSFAAL NQVATSGLSPPERRAAGQYQTVQLAAVVAVVAALLAAAEPPGATAEARGLSEA GADPGRA"		ORIGIN	
Query Match 8.0%; Score 347.2; DB 1; Length 5495;			
Best Local Similarity 56.7%; Pred. No. 2e-21;			
Matches 773; Conservative 0; Mismatches 563; Indels 28; Gaps 6;			
QY	2942	CGTGCCGAGCGCGACCGGGCGCGCTACGGGCGCGCGCGCGCGCGCTGAGTCTGCC	3001
DB	4201	CGAGCAGAGAGCCCCACAGCGGGGGCTTCAGAGCGCCCCCTCGCGCGCGAGCGGGCG	4142
QY	3002	CTGCGTACCGAAGCGTGGCGG----GTGAGAGAATCGTTCCGCCCTGTGGCAATCGACGTAC	3057
DB	4141	CGGGGTGCGTCTGTCACGGCGTCCACCGAGGAGGGTGC CGCGCTCGCGTCGACGGAAG	4082
QY	3058	TGGCGGTGATCCACCGTAGTCTCGAGGCGCAGAAAGGCCACACAGTCGCGCGATGTCG	3117
DB	4081	GAGCCGCGTGAACCGAGCGCGCTGCTCGAGACGAGGAAGGCGACACAGTCGCGCGACGTCC	4022
QY	3118	TGCGGTCTCGCGATCGCGTGTGAACACGAGTGTGGCGGCGAGTGC CGCGCGCGCTCGGGG	3177
DB	4021	TGCGGTTCGCGACCCGGTTGAAGTGAACAGGGCGCGCATCGCTCGACCGCTCGGGG	3962
QY	3178	GTCTGCCCGCGCGTGCCTCATGTTCGTCTCCAGAAACCGGCGCGCACCGCGTTGACC	3237
DB	3961	ATGTGCAACACGCGTTGCGGTTCGGGT-----GATGCCGGGGCGACGCTGTTGACG	3908
QY	3238	GTGATCCCCCGTTCCCGAGTTGCTTGCGCAGGGCGAGCGTGTGACGCTTCACCGCACCC	3297
DB	3907	GTGATGCGCGCGGGCCCGAGGGCCCTTGCGGTAGTGCAGCGCGAGCTGTTGACGCGCGCC	3848
QY	3298	TTGGTTCATCGCGTATCCGATGAGCTCGGGGAACGCGCGCGGTGCGCGCGACGAGATG	3357
DB	3847	TTGCTCATGGCGTAGGCGATCTCGTTCGGGGTTGGCGAACCGGGTTCAGGCCCGAGAGATG	3788
QY	3358	TTGATGATCCGCGCGCGCTCGCGAGTCTGTTTTCAGTCCGTGTGACGACGAGAACGCGGT	3417
DB	3787	TTACAGATCCGCGCGCGTTCGGGCATGTTGCTGACGGCGCGCTGCGACGAGGAACGCG	3728
QY	3418	GCCCCGAGCTTGACCGCGACAGTCTCGTTCGAGACCTCTCTCGGTGACTCTCGGTATCCGT	3477
DB	3727	GCCCCGGGCTTGACCGCGCAAGATCCCGGTTCGAACTGTCCGGCGTGTGCTTCGGGGCTTG	3668
QY	3478	CCCGAGCGCTGACCGCGCGTGTGTTTTCACGAGATGTTCGAACTCCGGGGCGCACTCCGAAC	3537

DB	3667	ACCCGCCCATACAGCCCGCGGCTTTGTTGACAGGATGTTTCACTCGTGGAGCCGCGTGGC	3608
QY	3538	TGCGCCATCCGGCGTTCGAACCGCGCTAGAGCGCGCGCGCTCACCAAGACCCGAGT	3597
DB	3607	TCCTTTCAGGCCCTGCTCCAGGGCGAGAAAGCTCTGTGCACGTTCGCGGGGCGCCGAGT	3548
QY	3598	TGCGCCCGGATGGCAACGCTGTCCCGCGCTGTCTCCGGATGGTCTTCGAGCGGTCTCTCGC	3657
DB	3547	TGCGCGGTAGCGGAACGCGCGCGCTCAGCTTCGATCCCGGCGACGACCTTCGCG	3488
QY	3658	GCGCCCGCTCGCTGTGCTGCTAGTGAATGTCACAGAGCGCGCTTCGCGCGCGCAGCGC	3717
DB	3487	GCGCGTCTCTGTTGGTGCCTAGTGCAGCGCGCTCAGCGCGCGCTTCGCGCGCGAGCGC	3428
QY	3718	AGGGCGATACCGGTTCGATGCCCGCGCTTCCCGCGTTCACAGGGGCGGTCTTCGCCCTCC	3777
DB	3427	TGCGCGGTGCGCGCGCTCAGCTTCGCTGCGCGCTGCGCGCTGCGAGCGCGCTCTCCCGGTC	3368
QY	3778	AG-----CGGTCTTCCATACCTCGTCCCATGTGTCACGCATATACAGCCCCCGCG	3826
DB	3367	AGTTTGCCCATCTTCGTCTCTGACGTGGTGTGTTGTTGCTGGGATGCGTGGGTGCGGT	3308
QY	3827	TGCGTGAAGC-GACCATGGGCGCGCTCGCGCGTTTCCGAATCGACGCTCAGACCTACCTG	3885
DB	3307	GCGGTGCGCGACCGCGCGCTGAACACAGCGCGCGAGCGCGCGTCTGAGTCTCTTCGCGGGTG	3248
QY	3886	TGACCGGTTCAGAGCGGGCGCGAGTGGCGCGTGTGAGCGCTGGGGCCAGATCGGGCGCG	3945
DB	3247	TGCGCGCGCCCGAGCGAGGTATCCGTTCGGGGGTACGAGAGGGCGCGCGCTTCGGGCGC	3188
QY	3946	GCGCAAGC---GGGAACCGCGCGCGTTCAGGGGTTCAGGGGTTCGCGGGAGCGCCAGGCC	4001
DB	3187	GCGCGCGCTCGCGCGGGCGCGCTACGCTGTTCGAGCGCACCGGACCGCGCGGGCG	3128
QY	4002	GGTTCAGGGCACCG--ACCGGATCGAGTTCGGGCGTGCACACGCGGCCACAGTCTTCGCGG	4059
DB	3127	ACCTTGGCGAACGCTCCGCGCGCTCGGTTCAGCACAGCAGCAGCGCGCGCGCTTCGCGG	3068
QY	4060	CCGAGCTTCGAGTCTGACGGGTACAGAGCGCGCTTCGCGCGAGTCTGAGTTCGACGCTGG	4119
DB	3067	AGCAGGTTCAGGGTGGAGCGGGTCCGCTGCGCGCCAGAGGGCGGGCTTCGGCAACCGG	3008
QY	4120	CCGCGCGGTGGGTGAGCGGTTTCGCGCAGGGGCGGAAGCGGGGAGGTTCGCGCGCGCAGC	4179
DB	3007	GTCCCGCGAGCTGGGTTCGCTGCGCGGTGCGTGGCGAGCTCCAGGCGCTGATCGT	2948
QY	4180	ATCATGGGCGGGCGCGGTGAAACCGCGCGCGCGCGGTTCGAGGGTCTCTCCAGGACG	4239
DB	2947	CCCGCAGGTGCGCGCGCACCGCTCTGACACCTTCGATTCGCGGATCACCGAGCGCAGC	2888
QY	4240	TCTAGACCGCGCAACCGCGCTTCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	4283
DB	2887	GCCTCGACCTCCCG	2844
RESULT 9			
SC0939127	290850 bp	DNA	linear BCT 16-APR-2005
LOCUS	Streptomyces coelicolor A3(2) complete genome; segment 24/29.		
DEFINITION	AL939127 AL023496 AL033861 AL031155 AL031182 AL031232 AL035161		
ACCESSION	AL035205 AL035206 AL132824 AL512902 AL590982 AL591083 AL591084		
AL645882			
AL939127.1	GI:24429552		
VERSION			
KEYWORDS	Streptomyces coelicolor A3(2)		
SOURCE	Streptomyces coelicolor A3(2)		
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
REFERENCE	1		
AUTHORS	Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,		

Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neil,S., Rabinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S., Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S., Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrell,B.G., Parkhill,J. and Hopwood,D.A.
Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)
Nature 417 (6885), 141-147 (2002)
12000953
2 (bases 1 to 290850)
Bentley,S.D.
Direct Submision
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sds@sanger.ac.uk
On or before Oct 30, 2002 this sequence version replaced
gi:20520661, gi:20520749, gi:20520754, gi:20520755, gi:20520688, gi:20520689, gi:20520815, gi:20520683, gi:20520917, gi:20520686, gi:20520864.
Location/Qualifiers
1..290850
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
97..1101
/gene="SCO6257"
/note="synonym: SCAH10.22"
97..1101
/gene="SCO6257"
/note="SCO6257"
binding lipoprotein, len: 334 aa; similar to TR:CA841563 (EMBL:AL049727) Streptomyces coelicolor putative secreted solute binding protein, 337 aa; fasta scores: opt: 1003 z-score: 1130.9 E(0); 49.8% identity in 325 aa overlap, z-score: 1130.9 E(0); 49.8% identity in 325 aa overlap, z-score: 1045.2 E(0); 49.8% identity in 325 aa overlap. Contains match to Pfam entry PF00532 Peripla_BP_like, Periplasmic binding proteins and LacI family and a match situated in the correct position to Prosite entry PS00013 Prokaryotic membrane lipoprotein lipid attachment site. Contains also possible N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=1
/product="putative ABC transport system sugar binding lipoprotein"
/protein_id="CAB60176.1"
/db_xref="GI:6273662"
/db_xref="InterPro:IPR001761"
/db_xref="UniProt/TREMBL:Q9RKT5"
/translation="MARFTWGTGIALAGSLVSLAGCSSTGKRAEDARKAASERGA AVDPRTWFTMTTHSGDGTFDWIVQSGAEQAVKDNINFLYSHDAEQOQALVNAALDKYVDGILVTLAKDPDKMSALABAHKAGIPVITVNSGSESKERFALTHVQSDETIA IGKAEELNERQKQAVCLVHQGNVGHQSCDGVKERTFDGKQRLVYNGTSPVQVS ATEAKLQTDSDVAVTGYGADYAVKAKQAGSKAEIDITFDLNAKVAAGLADGTLG FAVDQOPLYQGYEAVDLLWLYKYNADVLGGRGRLVLTGPQIITKDDAALADYTKRGTR"
133..165
/gene="SCO6257"
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"
241..966
/gene="SCO6257"
/note="Pfam match to entry PF00532 Peripla_BP_like, Periplasmic binding proteins and LacI family., score 13.80, E-value 2.1e-06"
1098..2138
/gene="SCO6258"

/note="synonym: SCAH10.23"
1098..2138
/gene="SCO6258"
/note="SCAH10.23, probable ABC transport system sugar permease, len: 346 aa; similar to TR:O68120 (EMBL:AF010496) Rhodobacter capsulatus ribose transport system permease protein RbsC; fasta scores: opt: 445 z-score: 508.0 E(0); 6.4e-21; 30.5% identity in 351 aa overlap and to SW:RBSCECOLI (EMBL:L10328) Escherichia coli ribose transport system permease protein RbsC, len: 321 aa; fasta scores: opt: 286 z-score: 329.9 E(0); 5.3e-11; 30.1% identity in 329 aa overlap. Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative ABC transport system sugar permease"
/protein_id="CAB60177.1"
/db_xref="GI:6273663"
/db_xref="GOA:Q9RKT4"
/db_xref="InterPro:IPR001851"
/db_xref="UniProt/TREMBL:Q9RKT4"
/translation="MNATQVRADEIRILQTSRLKRLGRPELGSVVGAIIVFPFAFF ADSFLHAASLTSLVYAASTIGAFNAPVALLMIGGFDSAGVMTVSSALVSMFSYQM TANVWGVVSLVTLAIGAFNGFMTLTKLPSTIITLGLTFLMLTGLTKLVDTG VSTKSIADMEGFPSSAODVFASITIGVGPKVTILMWLALVAVASWILRTLRAGNWIIF AVGNKQDARAVGYPVAKTKIGLYMGVGFAGWISQHLHLSFSDVVDVQSGEGVGNELIYI IAAVIGGLITGGTGSVAVGAVGAFIFGTSKGIIVFAEWNPDPWFKEFLGMLLLATLL NAWVRKRAEATK"
2124..2127
/gene="SCO6258"
2135..2926
/gene="SCO6259"
/note="synonym: SCAH10.24"
2135..2926
/gene="SCO6259"
/note="SCAH10.24, probable ABC sugar transport ATP binding protein, len: 263 aa; similar to TR:O68121 (EMBL:AF010496) Rhodobacter capsulatus ribose transport ATP-binding protein, 305 aa; fasta scores: opt: 612 z-score: 706.1 E(0); 5.9e-32; 45.6% identity in 217 aa overlap, to SW:RBSAECOLI (EMBL:M13169) Escherichia coli, ribose transport ATP-binding protein RbsA, 501 aa; fasta scores: opt: 509 z-score: 585.6 E(0); 3e-25; 31.5% identity in 254 aa overlap and to TR:CAAL5787 (EMBL:SC7B7) Streptomyces coelicolor St7B7.07, 260 aa; fasta scores: opt: 592 z-score: 521.2 E(0); 1e-23; 43.1% identity in 246 aa overlap. Contains pfam match to entry PF00005 ABC tran, ABC transporter and Prosite PS00017 ATP/GTP-binding site motif A (P-loop) and PS00211 ABC transporters family signature"
/codon_start=1
/transl_table=11
/product="probable ABC sugar transport ATP binding protein"
/protein_id="CAB60178.1"
/db_xref="GI:6273664"
/db_xref="GOA:Q9RKT3"
/db_xref="InterPro:IPR003439"
/db_xref="InterPro:IPR003593"
/db_xref="UniProt/TREMBL:Q9RKT3"
/translation="MTRNEDRTALVELSGVSKYGNVRALEGVSLVHAGEITCVLGD NGAGKSTLKIAGLHQHDDGGLSDGBETRLSSPREALDRGIATYQDLAVVPLMPV WRNFFLGSPPRGKGVAFKRMVDVHMRTTHAELLRMGIDLRDVPDIPGLSGERQCV AIARAVYFGAKVLVLDPEPTAALQKQSGVLKVAARDDQGLGVLTITNPHAVLVG DRFVLLKSGTVMVANTQTRDEVTLDELITQMGAGTGLDRLRHELRG"
2240..2815
/gene="SCO6259"
/note="Pfam match to entry PF00005 ABC tran, ABC transporter, score 168.70, E-value 9.6e-47"
2261..2284
/gene="SCO6259"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
2588..2632

misc_feature
misc_feature
misc_feature

gene	/gene="SCOG259"	
	/note="PS00211 ABC transporters family signature"	
CDS	3042..4190	
	/gene="SCOG260"	
	/note="synonym: SCAH10.25"	
	3042..4190	
	/gene="SCOG260"	
	/note="SCAH10.25, possible sugar kinase, len: 382 aa; similar to SW.GLK STRCO (EMBL:X65932) Streptomyces coelicolor glucokinase (EC 2.7.1.2) (glucose kinase) Glk, 317 aa; fasta scores: opt: 1022 z-score: 1142.9 E(): 0; 46.6% identity in 311 aa overlap and to TR.CAB51974 EMBL:SC6E10 Streptomyces coelicolor SC6E10.20c, 317 aa; fasta scores: opt: 1022 z-score: 1035.4 E(): 0; 46.6% identity in 311 aa overlap. Contains a match to Pfam entry PF00480 ROK, ROK family and Prosite entry PS01125 ROK family signature"	
	/codon_start=1	
	/translation="MSYTRDFTAPIGSRRAVLTGTRRSHLTAPRVTVCIDIG GTKWAGVVDADGNILEKLTETDPKSPKVEDTIVELVDLSDRHDVHVGIGAA GWDADNRNLFPAFLHSWRNEPLDRDTRAGRLAVPLVDNDTAANAEWRFGARGED HLMTITLTGTCGGLIDGQVKRGYGVAGFQHMVPGHRCPCGNRGWEQYSSG NALVREALEAADS PVAYGIIEHVKSIGDITGPMITELAREGDMCVLLODIGOW LGVGNLAALADSPCFVIGGVSAADLLIGPARDAPKRLTGRGYRPEARIVRAQL GPEAGMGAADLSLVARFRRAKRRRVERVRYEARRESRL"	
misc_feature	3159..3704	
	/gene="SCOG260"	
	/note="Pfam match to entry PF00480 ROK, ROK family, score 215.50, E-value 5.4e-63"	
	3540..3623	
	Query Match 7.7%; Score 335; DB 1; Length 290850;	
	Best Local Similarity 65.2%; Pred. No. 5.3e-21;	
	Matches 526; Conservative 0; Mismatches 275; Indels 6; Gaps 2;	
QY	2978	CGGCGGGCGGCGGTAGCTGCGCCCTGCGTACCGAAGCGTGGCGGTGAGAGATCGTTC 3037
DB	62506	CGGCGGTTCTTCGGCGGGCGGCGAGCCCCCGCGGAGCTCAGAGCGCGAGCGGAAC 62565
QY	3038	CGCTGTGGCATCGACGTACTGGCCGGTGCATCCACCGTGCAGTGCCTGGAGGCCAGAAAGG 3097
DB	62566	CGCCGCTGGCTCGATGTTCTGGCCCGTGCACCGCGGGCGTCTGTCGGAGCCAGGAAGG 62625
QY	3098	CCACACGTCGGCGGATGTCGTGGTCTGCGGATGCGGTGAAACCGAGTTGGCGGCCA 3157
DB	62626	CGACCACGTCGGCGACGTCGTGGGCTGGCGGCGTGGAGCTGAAGACCGAAGACGCGGC- 62684
QY	3158	GTGCGCGCGCCCTCGGGGGTCTGCGCGCGCGTGGTTCATGTCCTCTCCAGAAC 3217
DB	62685	--GCGTGGGCGCTTGGCGGGGTCGGCCATCCACGGGTGGATGTCGTCTCGATGGTGC 62742
QY	3218	CGGGCGCACCGCTTGACCGTGCATCCCGCTTCGCCCGTTCCTGGCCGAGGGCGAGCG 3277
DB	62743	CGGGGAGACGCGTGTGACGTGATGCCGCGGAGCCACAGGTCTGGCCAGGGTGGAGG 62802
QY	3278	TAAGCGTGTCCACCGACCCCTTGGTTCATCGGTATCGATGACGTCGGGGAAACGCGCGCC 3337
DB	62803	TCAGGACCTCGACCGCGCCCTTGGCGGGCGGCTAGGAGTTCATGCCCGGGAAGCGACCT 62862
QY	3338	GGTCCGCGCACACAGATGTTGATGATCCGCGCCCGCTCGCGAGTGTTCAGTCCGT 3397
DB	62863	TGGTGACACCGGAGGAGATGTTGACATACGCGCCCGCTGCGCGACGCTCCAGGCCCT 62922
QY	3398	GCTGGACACGAGAACAGCGGTGCGCGAGCGTTGACCGCGCACAGTCCGTGCGAAGACCTCCT 3457

Db	62923	TCTGGATGATGAAGAAGGGGCGCTTGGCGTTACCGCGAAACACCTTGTCTGTACTCTGGCGCT 62982
QY	3458	CGGTGACTTCCGTGATCCGTCGCCGAGCCGCTGACGCCCGGCTTGTTCACAGGATGTCCA 3517
Db	62983	CCTCGACTCTGTGATGAGGCCCGGGCGGGGATGCCCGGCTTGTTCACAGGATGTCCA 63042
QY	3518	ACTCGG---GCGGCACATCCGAATCGCCCATCCCGGCGTGAACCGCGCGTGAAGCGCGG 3574
Db	63043	GGCGGAGGACACCGTCCCGCTACTGACCGATCTCGCGGTGGAAGGCTGCCACAGCGCCT 63102
QY	3575	CCGCGTACCCACGACGCGGAGTTCGGCCCGGATGGCCACGCTTCGCGCGGTGCTCC 3634
Db	63103	CGGCGTCAACCGGACCCCGCTGCGAGCGGAAGGACGCGCGCCCGCGCCT 63162
QY	3635	GGATGCTCTGACGCGTCTCTCGCGCCCGCTCGCTGCTGCCGTAGTGGACTGCGACGA 3694
Db	63163	CGATCGAGTGAACCGTCTCTTGGCGGCGCTCTGCTACTGGATAGTTCAGGCGGACCA 63222
QY	3695	GCGCCCGCTTCGCGCGCCGACGCGGCGATACCGCGTTCGATCCCGCGCTTCCCCCGG 3754
Db	63223	GCGCGCGCTCTTGGCCGCGCTGGGCGATGGCAACGCCGATGCCCGGCTGCCACCGG 63282
QY	3755	TCACAGGCGCGCTTTCGCCCTCCAGCG 3781
Db	63283	TGACAGCGCGCTTCTGTCTGTGAGGG 63309
RESULT 10		
AY956334/c		
LOCUS	AY956334	17455 bp DNA linear BCT 04-AUG-2005
DEFINITION	Streptomyces aureofaciens strain CCM3239 auricin biosynthesis gene cluster, partial sequence.	
ACCESSION	AY956334	AY033654
VERSION	AY956334.1	GI:61968686
KEYWORDS	Streptomyces aureofaciens	
SOURCE	Streptomyces aureofaciens	
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	
REFERENCE	1 (bases 1 to 17455)	
AUTHORS	Novakova,R., Bistakova,J., Homerova,D., Rezuchova,B. and Kormanec,J.	
TITLE	Cloning and characterization of a polyketide synthase gene cluster involved in biosynthesis of a proposed angucycline-like polyketide auricin in Streptomyces aureofaciens CCM 3239	
JOURNAL	Gene 297 (1-2), 197-208 (2002)	
PUBMED	12384301	
REFERENCE	2 (bases 1 to 17455)	
AUTHORS	Novakova,R., Homerova,D., Feckova,L. and Kormanec,J.	
TITLE	Characterization of a regulatory gene essential for the production of the angucycline-like polyketide antibiotic auricin in Streptomyces aureofaciens CCM 3239	
JOURNAL	Microbiology (Reading, Engl.) 151 (8), 2693-2706 (2005)	
REFERENCE	3 (bases 1 to 17455)	
AUTHORS	Kormanec,J.	
TITLE	Direct Submission	
JOURNAL	Submitted (30-APR-2001) Gene Expression, Institute of Molecular Biology, Slovak Academy of Sciences, Dubravska 21, Bratislava 84251, Slovak Republic	
REFERENCE	4 (bases 1 to 17455)	
AUTHORS	Kormanec,J., Novakova,R., Homerova,D. and Feckova,L.	
TITLE	Direct Submission	
JOURNAL	Submitted (08-MAR-2005) Gene Expression, Institute of Molecular Biology, Centre of Excellence for Molecular Medicine, Slovak Academy of Sciences, Dubravska 21, Bratislava 84551, Slovak Republic	
REMARK	Sequence update by submitter	
COMMENT	On Apr 3, 2005 this sequence version replaced gi:14346013.	
FEATURES	Location/Qualifiers	
source	1..17455	
	/organism="Streptomyces aureofaciens"	
	/mol_type="genomic DNA"	
	/strain="CCM3239"	

```

gene
/db_xref="taxon:1894"
complement(416..1033)
/gene="aur1R"
CDS
complement(416..1033)
/gene="aur1B"
/note="putative transcriptional repressor"
/codon_start=1
/transl_table=11
/product="Aur1R"
/protein_id="Aur1R"
/db_xref="GI:61968687"
/translation="MIKQBRVTRRALIESAARVPGRGYAEATLSMISVGAGSPG
ALHEHPENKAATAAAVEAAARALSAAREVARTSAQALADSHALAGLLSDTV
ARAGQLSEPTPSYALPGWHEYHRLLAABAHEDALLPCLNRQSVAITVVAATL
GFEVLGRDDPQWLAPRTLAFGRVVMVPCLAAPATLFLGAAGRS"
1192..1197
1216..1221
1261..2031
/gene="aur1P"
1261..2031
/gene="aur1P"
/note="response regulator-like protein; transcriptional
activator essential for auricin biosynthesis"
/codon_start=1
/transl_table=11
/product="Aur1P"
/protein_id="Aur1P"
/db_xref="GI:61968688"
/translation="MNQRTMAVLEPHTVAAGNGCFSPAGERTPOHPVPAQSALWELL
IVESNSAEATLARGRRHGHEVDIVTGTGEALRVYAEADLVLLLELPDLOGLEVC
GIRAHEITPIVATGSELDRLVGLQAGADDYLAQPGFRELMAWMAEMRRARPR
PAFATITCGPLSDSTREVTLDGVEVTRKEFDMLYLLASHPDTVIPKMLMQQW
GDSWSRRVTDVTHVSLRNKIGAWIITRGVGRFG"
2099..3001
/gene="aur10"
2099..3001
/gene="aur10"
/codon_start=1
/transl_table=11
/product="aur10"
/protein_id="Aur10"
/db_xref="GI:62177082"
/translation="MTVTALCHGESCRGADATGTRRRRAADRGSLRCRCRHLAAGL
GEVAHYEBCRGVILGAGPGSLERTTGGESTGLPENGAAADVRAEMVSVLGSWGLV
AQOGLTAPRTVADULTAFLLRNADWLAHPAAGEVCDEVARLVRAGRRAAPPEPART
FRIGTCPEPCSCGALALSVRGAERDRAVCDADPAHRTSDRWSELSDSLESGRAP
GAVAGVPAPAPAPATERNLTAADIAQLRHTPTGTVYRLASEERWRRNRAGRIFY
AEGDVACFAAGRRPAPPPPHR"
3170..3175
3193..3198
3250..4716
/gene="aur1A"
3250..4716
/gene="aur1A"
/note="oxygenase-like protein"
/codon_start=1
/transl_table=11
/product="Aur1A"
/protein_id="Aur1A"
/db_xref="GI:61968689"
/translation="MDAPVTVVAGPAGMMLAGELRAGVAVIVLERLERRTGESRGL
GFTARTLEPQDQIGILRFGDIERSLTHGFGPLDFGLDGCQAARTVPSVTETH
LEWARTGLGADFERGHEVLSVDGDTVAVEVRPGGTHLTRADMLVCGDGGRSVTRK
AAGDPFGTAATMEMLADIKGIELKPMIGETLPGMWVWGPLPGSTTRIVVCERG
PPRRRAPSPFEVAAAWKLTGEDI SAEPVWVSFAFGDARQVTEYRGRVLLAGDA
ATHLPAGGQWNTGIQDANVLGWKLASVVLGRAPLALLDTHYGERHPVGRRLMMNR
AQGLFLSGAEVQLRSLRFLMAYEDVARHAAVMTGLEITTYDVKGKGPLLGRMP
GLELAGFGRTSAAALCPARGVLLDLANPLRQRQAAMGDRVDTVTARPAGAGGE
LPDITTAALIRPDGYVAAWAPGSHDLPALERWFGPA"
4891..5538
/gene="aur1B"
4891..5538
/gene="aur1B"

gene
/note="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="Aur1B"
/protein_id="Aur1B"
/db_xref="GI:61968690"
/translation="NATPLATIERAPSGPAGERREIVRAARTVLAREGTRCTABATA
AEAGVTADTVLHFFDHEQLLGLVLLDSSASVAALAAVADSHLEEVTLDEADLIAG
RAWLAPAEAFREHFVIRHISAEVTRLPTEAVDKWQATGTPQARRELARLRRLASMG
LLDISDAQAEARERFILLVPSGVVQSRFLGALFLRDEADALLADGVADFIRLYRP"
5621..5953
/gene="aur1C"
5621..5953
/gene="aur1C"
/note="cyclase-like protein"
/codon_start=1
/transl_table=11
/product="Aur1C"
/protein_id="Aur1C"
/db_xref="GI:61968691"
/translation="MTHSNLIVARMSMSSDQVAKLFTFEDKTDMPGRMGTLRRRLFM
YKGLYFHLQDFAEADDGQRIQESRTDPRFMRISHDLLPFIQPYDPATWRTPADAMATR
FYDWGASE"
5991..7259
/gene="aur1D"
5991..7259
/gene="aur1D"
/note="B-ketoacyl-ACP synthase-like protein"
/codon_start=1
/transl_table=11
/product="Aur1D"
/protein_id="Aur1D"
/db_xref="GI:61968692"
/translation="MTRRVITGIGVLAPGGITTAFWNLSSGRTATRGITLFDPA
FRSVAEVDHPERHLSLEVRMDRAAQFQVWATREALADSGLDLAAFDYRIGI
AMGTGVAISGLDTRVAVASDEGRVIAQVDHVPAGHLYDHFVSPSPATEVAVAGAE
PSTVSTGCTSGIDSVAHAVDLIRREGSADIMLTGATDAPISPLTACFDALKATTALN
BEPFASRPFDAITRKGFVLGEGAAFPVLEELGAARRGRGHVYAEIAGYASRSASFMT
GLRPDRENAEAITVALDEARLDPGDLDVYNAHSGTRQNDLHETHTAAYKLALGPHHR
TPVSVKSMIGHISLGSVSGMEIAASVLAHMEHVVPPTANLRTDPDCELDLYVPLVARD
HRTDAVLTVSGFGGFSQAMWLARPDRTA"
7256..8470
/gene="aur1E"
7256..8470
/gene="aur1E"
/note="chain length factor-like protein"
/codon_start=1
/transl_table=11
/product="Aur1E"
/protein_id="Aur1E"
/db_xref="GI:61968693"
/translation="MSARILVTGIGVAAPSGLGVDFWSVTRIGKNAIGPVTRFDASA
YPSRLAGEITHGPEKHELPQLVPODRTVLTALVAADCAFADAGIEPGTIDPYAMGV
VTAAGAGGFEEAENELKWSGAKHVSVAQSPFVAVNSGOISIRNGLRGPAGVVI
SDQAGDLAQAARQLRKSGKLIATGFDAPICSLGWASHLGGMLSTSDPERAYL
PFDDAAAGVYPGEGGAMLILEDEDSARDRGARTVYGFAGYGTALDPKPGSGREPLR
RAIDVALTDAACHPAEVDVFDAGRHPDLRDEAEAITRVFGPRVAVPTVPTKMTGRI
NSGGAPIDVVSVALSNREGLIPTTNVELSDAYDLDLVAVRPRTASVRTALVLARGG
GFNSAVVVRVD"
8568..8852
/gene="aur1F"
8568..8852
/gene="aur1F"
/note="acyl carrier-like protein"
/codon_start=1
/transl_table=11
/product="Aur1F"
/protein_id="Aur1F"
/db_xref="GI:61968694"
/translation="MKGSIISAPEFTDDIKRILLEGAGDEGVLDGDLDSSEFETL
GYESLALLETGGRIERFEGITLDDVDFTDNTPRSLVTGINESLSALVDA"
8946..9731
/gene="aur1G"

```

Db	12352	CGGCCCATGCCCCGGGCTCGATCCGGTGTACAGCGCGGTCCGGCCCGGAGCGCGTCCGGTCCGCTG	12293	
QY	3790	TACTCTGTCCTCATGTGTACGCAATATCAGCCCCCGCGTGTGAGCAGCACCACCTATGCGCGC	3849	
Db	12292	CCGGCGCCCCGGACCTCCCGGGCGCGCTCACCGGGTGTACCCGCAACCGCTTCGCGCGGG	12233	
QY	3850	GCTCGGCGGTTCGAATCGACGGTCAACGCTTACTGTGACCCGCTCAGACGGGCGCGGAG	3909	
Db	12232	GCGGCTGCGCGGGGGAGCGCAGCGCGCGCGCCCGAACACCTTCCAGGGCGCGCGCTC	12173	
QY	3910	TGGCCCGGTTGTACGCGTGGGGCCAGATCGGGCGCGCGCACCGGGAAACCGCGCGCGCTC	3969	
Db	12172	AGTCTCTCCCGGGGCGCGTGTCCGCGTCCGCTCGACCGACGACGTCGCGCTCGGGA	12113	
QY	3970	AGGGGTTCAGGGTTCGCGCGGACCGCCAGGCGCGTTCAGGGCACCGACCGGATCGAGGTG	4029	
Db	12112	CGCACGAGGAGGCGCGGGCGCGCTCCCGCTCCGCTCCGCGCGCGTGGCGATCCGT	12053	
QY	4030	GGGTGCGCACGGGGCCACAGTCTCTCGGGGCCAGTCCGATCTGTAACGCTACACAGAC	4089	
Db	12052	ACGCGGTCCCGCAGGGCGCGCGCGCGCGCTTCCTCGCGCGTCCCTCGTGCGCC	11993	
QY	4090	CCGCTCCGCGCAGTCTGAGCTGAGCTGGACGTGCGCGCGCGGTGGTGGGCGGTTCGCCAC	4149	
Db	11992	GGGGCAGCAGCAGCAGTCCCGCGCGCGCGCGCGCTCGAAGACGCTCGAGTGGTGGCGCC	11933	
QY	4150	GGCGGAAAGCGGGAGGTTCGCGCGCGAGCATCATGGGCGCGCGCTCGAAGACGCGCG	4209	
Db	11932	TCTTGGCCACCGCGCGCTTCGCGGACCGCGCGCGCGCGCGTCTCGGGAGCG	11873	
QY	4210	GC 4211		
Db	11872	GC 11871		
RESULT 11	AY034378/c	16643 bp	DNA	linear
LOCUS	Streptomyces sp. PGA64	silent angucycline-type gene cluster,		
DEFINITION	partial sequence.			
ACCESSION	AY034378			
VERSION	AY034378.1	GI:14280338		
KEYWORDS	Streptomyces sp. PGA64			
SOURCE	Streptomyces sp. PGA64			
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.			
REFERENCE	1 (bases 1 to 16643)			
AUTHORS	Metsa-Ketela, M., Palmu, K., Kunnari, T., Ylihonko, K. and Mantsala, P.			
TITLE	Engineering Anthracycline Biosynthesis toward Angucyclines			
JOURNAL	Antimicrob. Agents Chemother. 47 (4), 1291-1296 (2003)			
PUBMED	12654660			
REFERENCE	2 (bases 1 to 16643)			
AUTHORS	Metsa-Ketela, M., Kantola, J. and Ylihonko, K.			
TITLE	Cloning and Characterization of a Silent Angucycline-type Gene			
JOURNAL	Cluster from a Rubromycin B Producing Streptomyces sp. PGA64			
PUBMED	Unpublished			
REFERENCE	3 (bases 1 to 16643)			
AUTHORS	Metsa-Ketela, M., Kantola, J. and Ylihonko, K.			
TITLE	Direct Submission			
JOURNAL	Submitted (11-MAY-2001) Molecular and Cell Biology, Gallilaeus Oy,			
PUBMED	Kairiskulmantie 10, Kaarina FIN-20781, Finland			
FEATURES	Location/Qualifiers			
source	1. .16643			
	/organism="Streptomyces sp. PGA64"			
	/mol_type="genomic DNA"			
	/strain="PGA64"			
	/db_xref="taxon:161235"			
gene	complement(<1. .751)			
	/gene="pgak"			
CDS	complement(<1. .751)			
	/gene="pgak"			
	/note="similar to ORF12 from the tylosin biosynthesis"			

SVVRLRGFGFPETGEGRELLRADVAGIDVNNRRFELDCGLAI VARRPPDGVTRVMVH	
EFGAVPRGGRPEFDVVDARFKVTGDDISGTFPLWNSFHDASRQADSRYDHRFFLAG	
DAAHQMPICGGALNGLQDAVNLGNLAAQVAGRGPDGLDLSIHRHRAVGQVLSN	
IRTQALLLGSAAEVPYQVLAELIANSENVRSLHAGVSLDRIYDVGPGEPLLR	
RLPHWQLTDEGAVTSTAALRPGRVLLFPDGPQARHBEELSAAPAFADRVRTV	
TAGSCGTGGTDPGATVLRPDGVVWADPTGDLQALRRWFAPATAAAATAPAPAA	
PSTAARANNSSATATAKKWRTHMNRITSKALTVSGSRMGRAAAIRLAREGALVAV	
HYTARKDAEDVVAIEIKDGRAPTVRAELGVDPDVHELFLGLELREFTSGTDLDI	
LVNNAGVMGVKPEDITPEKPELFAVNAKAPFLIQRALKNMPDGGRIINISGLTR	
VANPDEIAYAMTKGANDQALHFAKYLGPRNITVNSVAPGITRNDNPVFGIPBAVEOM	
AQLSAFNVRGEPEDVADVAFATDEARWITGSPVDATGTTLLG"	
gene	11068..12534
CDS	/gene="pgaj"
	/gene="pgau"
	/codon_start=1
	/transl_table=11
	/product="putative transporter Pgaj"
Query Match 7.4%; Score 322.6; DB 1; Length 16643;	
Best Local Similarity 56.0%; Pred. No. 1.7e-19;	
Matches 679; Conservative 0; Mismatches 519; Indels 15; Gaps 3;	
Qy	2564 GCCCCTACGGCCCGCGGGCCCGCTAGTCTGCGCTCGTACCGAAGCGTGGCGG 3023
Db	11004 GCGCCCGCGCGGTACCGACGACCGGGGACCTGGCGTGGTGTCTGAGGGGCGGTCA 10945
Qy	3024 TCAGAGAAATCGTTCGCGCTGTGGCATCGACCTACTGCGCGGTGATCCACCGTGAAGTCGTC 3083
Db	10944 GCCAGCAGGTGCTCGGTGCGTTCGACGACGAGGACCGGTGATCCAGGGGCGCTCGTC 10885
Qy	3084 GGAGGCCAGAAAGCCACCGTCGGGATGTCGTCGGGTCTGCGGATGCGGATGCGGTTGAACAC 3143
Db	10884 GGTGCGAGGAAGCCACACCGTCCGACCTCTCCGCTTCGCGGACGCGGCTTGAAGGC 10825
Qy	3144 GGAGTTGGCGCCATGTCGCGGGCGCTCGGGGTCTCGCGCGCGCTGCTCATGTC 3203
Db	10824 GGACAGTGGCCCATCTGCTCCACCGCTCGGGAT-----GCGAAGACGGGTGTGTC 10771
Qy	3204 CGTCTCACAGAAACCGCGCGCACCGCGTTGACCGTGTATCCCGCTTCCCGAGTTGCT 3263
Db	10770 GTTGGGTGATGTCGGGCGCACCGAGTTGACCGTGTATTCGCGGGCCCGAGTACTT 10711
Qy	3264 GGCAGGGCGAGCGTGAGCGTGTCCACCGACCTTGTGTATCGGTATCGATCGACTC 3323
Db	10710 CGCGAAGTGCAGGGCGAGCTGTTCGACGGCGCCCTTGGTATGCGGTACGCGATCTCGTC 10651
Qy	3324 GGGGAACGCGCGCGGTTCGGCGAGACGAGATGTTGATGATCGCCCGCGTCCGCGAG 3383
Db	10650 CGGGTTGGCGACCCGGGTGAGCGCGGAGTGTATGATGATGCGCGCCCGCTCGGCAT 10591
Qy	3384 TCGTTTCAGTCCGTCTCGAACAGAAACAGCGGTGCGCGGACGTTGACGGCGACAGTCG 3443
Db	10590 GTTCTTACGGCGCGCTGATGAGGAAGAACCGCGCTTGGGTGACCGCGAAGATTC 10531
Qy	3444 GTCGAAGACCTCTCTCGGTGACTTCGTGATCCGTCCGAGCGCGTGAACCGCGGTGTT 3503
Db	10530 GTCGAACCTCTCTCGGGTGGTGTCTTCGGGTGACGCGCGCCCATGACACCGCGCTTGT 10471
Qy	3504 CACGAGATGTCGACTCGGGGCGACCTCCGAACCTCGCCCATCCCGGGGTGMAACGCGCG 3563
Db	10470 GACGAGGATGTCAGGTGCGGTGAGCCGGTGCCTCGCGAGCGCTTCTCAGGCCCGAG 10411
Qy	3564 GTAGAGCGCGCGCGGTCAACCAACGACCGAGTTTCGGCCCGGATGGCCAAACGCTGTCC 3623
Db	10410 GAAAGATTCTGTCAGCTCGCCCGGACCGCGAGTTTCGGCGGTACGGTGAAGGACGCGCC 10351
Qy	3624 GCCGTGCTCCGGATGGTTCGAACGGTCTCTCGCGCGCGCGCTCTCGTGTCTGCCGTAGTG 3683
Db	10350 ACCGTCTTCTCGATCTCGGCGACGACGCTCTCTCGCGCGCGCTCTTTCGCGCGGTGTAGTG 10291
Qy	3684 GACTGCCAGGAGCGCCCGTCCGCGCGCGAGCGGAGGATACCGGTTCGATGCCCGCG 3743

Db	10290 GACCGCGACGAGTGGCGCCCTCCCGGCGAGGCGTATGTCGCGGACGCCCGCCATGCCCT 10231
Qy	3744 GCTTCCCGCGGTACACAGGGCGGTCTTCCCTCCAGCGGTCTTCCATACCTCGTCCCATG 3803
Db	10230 GCTGGAGCCCGGTACACAGCGCGGTCTTCTCGTGAAGCTGTTCTATGTGTTCTCCACTT 10171
Qy	3804 TGACAGCATATCAGCCCCCGCTGGGTGAGCGACCCCATGCGGCGCGCTCGGCGGTTCGA 3863
Db	10170 CTTCCGCGTCCGTGTGGCCGACGAGTTGTTGTTGCGCCCTCGC-----CGCGGTGACGGT 10116
Qy	3864 ATCAGCGTTCACAGCTTACTCTGTGACCGCGTTCAGACGGGGCCGAGTGGCGCGGTTCGAC 3923
Db	10115 GCGCGCGGGCGGCGCGGTGGCTGCGCGCGCGCGCGCTTTCGAAACACAGCGCGCG 10056
Qy	3924 GGTGGGGCCAGATCGGGCGCGCGCACGGGGAACCGCGCGCGTTCAGGGGTTCAGGGGTC 3983
Db	10055 AGC-----GCCTGTTCCAGTCCGTGCGCGGGGTGTCGCGCCAGACGATATCCGTGGG 10000
Qy	3984 GCGGGACACCGCCCGGTTCAGGGCACCGATCGAGTTCGAGTTCGGCGGTTCGCACCGG 4043
Db	9999 CCGGACAGAGGGCGCGGTGTCGGGTGTCGCGCGCGGTGTCACGAGCGCGGTGAC 9940
Qy	4044 CCACAGTCTCTCGGCGCCAGCTCCGACTCGTACCGGTACAGAGCCCGGTCCGGCCGAG 4103
Db	9939 GGTGCGCACCGGTTCGCGCCAGGGGCGCGCGGAGCGAGTTCCTCTGTCGCGGGCTG 9880
Qy	4104 TCTGAGTGCACCTGCGCGCGCGGTGGGTGAGCGGTTCGCCAGGGCGGAGGGCGG 4163
Db	9879 CTGCGCGGGTCTGCGCGCGGACGAGGAGACCGCGCGCGCGCGCGCGCGCGCGCG 9820
Qy	4164 GAGTCCGCGCGG 4176
Db	9819 CGTCTGTTGACG 9807
RESULT 12	
AF322256/c 39428 bp DNA linear BCT 29-OCT-2002	
LOCUS Streptomyces antibioticus simocyclinone biosynthetic gene cluster,	
DEFINITION partial sequence.	
ACCESSION AF322256	
VERSION AF322256.1 GI:16223998	
KEYWORDS Streptomyces antibioticus	
SOURCE Streptomyces antibioticus	
ORGANISM Streptomyces antibioticus	
REFERENCE 1 (bases 1 to 39428)	
AUTHORS Galm,U., Schimana,J., Fiedler,H.P., Schmidt,J., Li,S.M. and Heide,L.	
TITLE Cloning and analysis of the simocyclinone biosynthetic gene cluster of Streptomyces antibioticus Tu 6040	
JOURNAL Arch. Microbiol. 178 (2), 102-114 (2002)	
PUBMED 12115055	
REFERENCE 2 (bases 1 to 39428)	
AUTHORS Galm,U., Li,S.-M., Schimana,J., Fiedler,H.-P. and Heide,L.	
TITLE Direct Submission	
JOURNAL Submitted (17-NOV-2000) Pharmazeutische Biologie, Pharmazeutisches Institut, Mikrobiologie/Biotechnologie, Universitaet Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany	
FEATURES	
source	
1..39428	
/organism="Streptomyces antibioticus"	
/mol_type="genomic DNA"	
/strain="Tu 6040"	
/db_xref="taxon:1890"	
<1..168	
/gene="sim2"	
<1..168	
/gene="sim2"	
/note="ketosynthase"	
/codon_start=1	
/transl_table=11	
gene	
CDS	

```
/product="Sim2"
/protein_id="AAL15580.1"
/db_xref="GI:16223999"
/translation="IPPTANLHDPEDCDLVVPTAREARLDRVLSVSGFGCFQSA
NVLTRPENTTA"
165..1379
/gene="sim3"
CDS
165..1379
/gene="sim3"
/feature="chain length determinant"
/codon_start=1
/transl_table=11
/product="Sim3"
/protein_id="AAL15581.1"
/db_xref="GI:16224000"
/translation="MTTSVVTGLGVAAPNGLGTADYWAATREGSGIGRVRFPDPSQ
YPSLAGVFGFVAEDHLPSRLLPQTHMTRLALVSADWALQDAGIRPEELPEVAAGV
VTASAGGFEGNELRALWKSQSYSAYSQSPAFYAVNTGQISIRHGLRPGSGVV
SDQAGLDALQARRQIRGSGQIVSGGIDASICSWGGAQTSGRMSTDDDPATYR
PFDGAAGHVPCEGALVLEEGAAAGAAAGVYGRYAGYGTTFDPKPGSGREPLRR
AIELATDAGLAPADIVVFADAAAVPDLRIEARALVEVFGPLPTAPKSMTGRL
NSGAGSLDVAALLAIRDGVTPPTINVTQADDYELDLVTAQARSARLSALVWARGH
GFNSALVQDAA"
1428..1694
/gene="sim4"
CDS
1428..1694
/gene="sim4"
/feature="acyl carrier protein"
/codon_start=1
/transl_table=11
/product="Sim4"
/protein_id="AAL15582.1"
/db_xref="GI:16224001"
/translation="MQTTEPTLDDLRLRDGAGASEVDVLEGGDIIDVAFEDLGYESI
ALLETCSRIRERYGISLDDDAVSEAPTPRALVAIVNTHLIAVAG"
1714..2499
/gene="sim5"
CDS
1714..2499
/gene="sim5"
/feature="ketoreductase"
/codon_start=1
/transl_table=11
/product="Sim5"
/protein_id="AAL15583.1"
/db_xref="GI:16224002"
/translation="NPGQKQVALVTGATSGIGLAVRALAQGHQVFCARNAKVA
VTIKELRGGLVDGTSCDVRSDVDKAFVAAAVRFGPDVVLNNAAGSGGVTAEL
TDELWADVIETNLTVFMVTVREVLTAGGRKSSGRIVNIASTAGKQGVVLGAPYAS
KHGVVGYFTKALGNELAPTGITVNVCPGYVTPMABRVROGYAAAYGTSEAILBKFO
AKIPLGYSTPEEVAGLVGLVSDTAASITAQALNVCGGLGNF"
2525..3472
/gene="sim6"
CDS
2525..3472
/gene="sim6"
/feature="cyclase/dehydrase"
/codon_start=1
/transl_table=11
/product="Sim6"
/protein_id="AAL15584.1"
/db_xref="GI:16224003"
/translation="MSDGSISHVEHTIEVDAPATDYYRLLAEVWNPRIFGPTIIVDH
VERGARTERIRIWTANGAAKSWTSRELDPEALRIDFQEVSAAPPVAMGGAWIIEP
VSEGACRVLHLDHYRALDAPEDLAWIEQAVDRNSRSELAALKANAEEVVVGADELML
SFQDSVRIDGSAKDVFDFLEADRWTERLPHEVAEVLTEQTPLQLVLTWETRAKDGST
HTTSSVRVCFPHERIVYQTVLPALMTLHTGYLLEEDGEGTVATSQHTVITANIR
AILGSGAGVAEARYVQVTALSTNSRATLGHAKDYAERD"
3478..4962
/gene="sim7"
CDS
3478..4962
/gene="sim7"
/feature="oxygenase"
/codon_start=1
/transl_table=11
```

```
/product="Sim7"
/protein_id="AAL15585.1"
/db_xref="GI:16224004"
/translation="MTGPTDVLVIGAGPVGLMLAGELLGLOGRVVVAERLTETPTES
RASTLHARTMEILQDRLGRLGEVPGDGTGHFGGIPRLDQQQSPYAGLMKVPQSRV
BEILGAWAIGLGAEEVRRGRELTLGLREDGAVVEAELSCDGTRELLRARYVVGDCGENS
TVRLAGPEFAGSAASRELIRADVAGIDVNNRRPERLPGLAIAARRPDGITRVWVHE
FGRTATPQAPDPDFDEHARTWQRTVEDISHTPTWINSFGDASQVTEYRGRILLA
DIAAQAQMPVGGQALNLGLQAFNLGWLAAABICWNAFPGGLDLSYHERHAGSVRLS
DIAAQAQMPVGGQALNLGLQAFNLGWLAAABICWNAFPGGLDLSYHERHAGSVRLS
RLPDAALDIEAGAATAWALLRAGTGLLDVAPEAPGHVDLRAVADGWADRVRVVAARP
RAGAVATPTGAVLVRPDGVHVVWADGGADRLRAALVQWFGSADRT"
5041..5805
/gene="sim8"
CDS
5041..5805
/gene="sim8"
/feature="oxidoreductase"
/codon_start=1
/transl_table=11
/product="Sim8"
/protein_id="AAL15586.1"
/db_xref="GI:16224005"
/translation="MKRLAGKTALVTGSSRGIGRATRLRLAREGATVAVHYADNADA
KATLAAITREGGRAPFLVRAELGVPDVTDLFQGLTAGREATGKEELNILLNAGVMG
GVKPEDLTPEJFDRLIAVNAKAPLFTQRLVLSIMPDGRIINIGTGLLRYPANPDEVAY
SMKSGALDQISILHFAALLGPRIITVNTVSPGTTNGSPIFDVPEAREFMANNAFKRI
AEPEDVDGVAFPLASDADDDARWITGTFTVDASGGTLLGG"
5936..6508
/gene="sim9"
CDS
5936..6508
/gene="sim9"
/feature="reductase"
/codon_start=1
/transl_table=11
/product="Sim9"
/protein_id="AAL15587.1"
/db_xref="GI:16224006"
/translation="MSSARLRVLAVLSTREGRLCPTVANWFAGQGRHQGGFDLLDLD
PYGLDIPVAPSPPEPQVVEIGQRLAADAFAVITPEYNHSPSAKLSDIWHYQMR
AKPVGFVSYGGLGGLRAVEQLRVFAELHAATVRDVTVSFHNAQLFDEAGDPRDEG
CNTAAKTMLDOLHWMGRALRQARLERPYQG"
6516..7202
/gene="sim10"
CDS
6516..7202
/gene="sim10"
/feature="putative phosphopantetheinyl transferase"
/codon_start=1
/transl_table=11
/product="Sim10"
/protein_id="AAL15588.1"
/db_xref="GI:16224007"
/translation="MARTESVQPAEAVQLWLRPEFVRPDELAALDSEERRRABAFPG
SADRMLYLSAHLARLRLAARLHVAPQEVIRVDRNGRPTLPGRPPHFSLSHSAG
LALLCTAPVIGVDVQRTLSRTTADLCGRHLHPASOELASVQPSARAHAFTRLWTRK
EAYLKGGLGVGLSRSLAADYLVGRDPTAGWTVDLPVPCGESHLSASAQVSRSPPLVTL
RWLLRSADHV"
7195..8781
/gene="sim11"
CDS
7195..8781
/gene="sim11"
/feature="carboxyl transferase/decarboxylase"
/codon_start=1
/transl_table=11
/product="Sim11"
/protein_id="AAL15589.1"
/db_xref="GI:16224008"
/translation="MSDVPSIPADPAPTRCRLEELRALRQAVADGPDPRATEAQHAK
GKLTARERVELLDGSGHEVRLRHRRAQGFGLGEARKPHDTGVVVGWGTVGRITVFV
YAHDFRIFGGALGEAQAOKIHKIMDAIAGAPLVSLNDGAGARIOGTPTALAGYGI
FORNTRASGVIPOISVMLGPCAGGAAYSALTDFVFWRETQSOMFITGDPDVQVATGE
KLSQGLGADYHSAVSGVSHFYDDQSCIEBVRVLLSLLPSNNREMPVQLSDDPV
DRNDSLLDLVPDAGNRKVIIEEVVDYGEYLEVYHERWATNVLCLVARIDGHVTG
IITANQPSLAGLVDISASEKAARFVQMCDAFNILPLVTLDDVPGLFPGVDQSHGGIIRH
```

	GAKLLVAYCNATVPRLISLIIRKAYGGYIVMDSRSIGADLAYAWPTNTEIAMVGEGAL NVIFRDIAAABDPDMRTQKIKEYKAELMHPTAAERGLVDVIDPAETRQVLTSAL AMLRTHADLPFSRKHGPPM"
gene	8796..9014
CDS	/gene="siml2" 8796..9014 /gene="siml2" /note="hypothetical 7.6 kDa protein" /codon_start=1 /evidence=not experimental /transl_table=11 /product="Siml2" /protein_id="AALI5590.1" /db_xref="GI:16224009" /translation="MTSDAMTDAIRVESGHLEPEELAALTWLLARIRGAAPDRAAGD GSLIVAATWRELFRANGFQSPSSWRAC" 9056..9829
gene	/gene=".siml3"
Query Match	7.3%; Score 316; DB 1; Length 39428;
Best Local Similarity	62.1%; Pred. No. 4.6e-19;
Matches 518; Conservative	0; Mismatches 310; Indels 6; Gaps 3; L1; L2;
Qy	2977 CCGGCGCGCGCCGCTAGGTTCGCCCTGCCTGCGTACCAGAACGTGGCGGGTCAGAGAATCGTT 3036
Db	5846 CAGGTCCGCGGCCCGTCCGATCGCGCGGGCCCAGCGCGGTTCAGCGCCCGCAGCAGCGTG 5787
Qy	3037 CCGCTCTGCATCAGCATCTGCCCGGTGATCCACCTGAGTCTCGGAGGCCAGAAAG 3096
Db	5786 CCGCGCTGCGCTCGCAAGAGGTCCCGGTGATCCAGCGGGCGTCTCGGAGGCGAGGAG 5727
Qy	3097 GCCACACGTCGGCGATGTCTGCGGGTCTGCCGATGCGGTTGAACACGAGTTGGCGGCC 3156
Db	5726 GCCACACGTCGCGCAGTCTCTCGGGTCTGGCGATCCGCTTGAAGCGTTCAITGTTGCC 5667
Qy	3157 AGTCCCGCGCGCCTCTCGGGGTCTGCCGCGCGTGGCTCATGTCCGTCTCCACGAAA 3216
Db	5666 ATGAATCTCCGCGCTCTGGGACCTCGAAGATCGGGTGCCTGTTTGGT-----GATG 5613
Qy	3217 CCGGCGCGCACCGGTGTACCGTGATCCCCGTTCGCCCGATTGCCCTGGCCAGGCGCAGC 3276
Db	5612 CCGGCGCACACCGTGTGACGGTGATGTGTCGCGGCGCGAGCAGCGGGCGAAGTGCAGG 5553
Qy	3277 GTAGCGTGTCCACCGCACCTTGGTTCATCGCGTATCCGATGGACTCGGGGAAACGCGCGC 3336
Db	5552 CTGATCTGTTGACGGCGCCCTTGCTCATCGAGTAGGCGACCTCGTTCGGGTTGGCGTAA 5493
Qy	3337 CGGTCGCGGCAGACGAGATGTTGATGATCCGCCCGCTGCGCGCAGTCTGTTTCAGTCGG 3396
Db	5492 CGCAGCAGTCCGGTACCAGTGTTCATGATACGTCCGCGCTCCGCGCATGATCGACAGCACG 5433
Qy	3397 TGTGAGACAGAGAACAGCGGTGCCGAGAGTGTGACGGCGACCACTCGGTGGAAGACCTCC 3456
Db	5432 CGCTGGGTCAGGAAGCGGAGCCTTGCGCTGTGACCGCGATGAGCCGGTCCAAGATCTCC 5373
Qy	3457 TCGETGACTTCGCTGATCCGTCGCGAGCCGCTGACGCCGCGGTGTTTACACAGGATGTCG 3516
Db	5372 GGGGTGAGGTCCTCGGGTTCACGCCGCCCATCAACCGGGCTTGTTTCAACAGGATGTTG 5313
Qy	3517 AACTCGGGCGGCACCTCCGAACTCCGCCCATCCCGCGTCGAAACGCCCGCTAGAGCGCGGCC 3576
Db	5312 AGCTCCTCTTCCCGGTGGCCTTCGCGCAGACTCTCGGTTCAGACCTCGAAGAGGGTGTGCG 5253
Qy	3577 GCGTCACCGACGACCGCGAGTTCCGCCCGGATGCGCCAACGCTGTCCGCCCTGCTCCGG 3636
Db	5252 ACGTCAACCGGCACACCCAGCTCCGCCCGGACCAAGGAAGGCGCGGCCCGCTCTTTCCTCG 5193
Qy	3637 ATGCTCTCGAGGCTCTCTCGCGCGCGCCCTCTGCTGTGCGTAGTGGACTGCCACGACGACG 3696
Db	5192 ATCGGGCGAGTGTCCTTGCGCGCTGCGCGTGTGTTCGGGCTAGTGAACGCGACAGTGC 5133
Qy	3697 GCCCGGTCCGCGGCACCGCGCAGGGCGCATACCGGTCCTGATGCCCGCCGCTTCCCGGTC 3756

gene	/translation="MHSTLIIVARMSPRDDQDVARLPSFEDRTTEMPHLMGTRRRQLFSY RGLYPLHQDPDEHGEAIEAAKSPSFVRISEDLPFIQAYDPTWRSADAMARRF YQWSAQ" /db_xref="GI:12744824.1" /gene="simA1" /db_xref="GI:12744824.1" /gene="simA1" /note="involved in polyketide formation" /codon_start=1 /transl_table=11 /product="putative ketosynthase SimA1" /protein_id="AAK06784.1" /db_xref="GI:12744820" /translation="MRRRVITGVGVGMAPGGIGVKNFWSLLSEGTATRGITFFDPAP FRSVAEIDPFPEKHGLGQPKIRMDRAAQLAVIATREAVADSGLDLADTPHRRGV TVGSAGVATGLDQBYRVVDDRLGLVDHRYAPPHLYCHFVPSFAAEVAWASGSEG PSTVVTGCTSLDLSHGAVALIREGSDVMITAGTADAPISITWACFDATKATTPRN DDPAHASPPDRSRNGFVLGECHAVLVEELGAOORRGARIYAETGGVASCNAFHMT GLPGRREMAEIRVALDEARLNPDIDITINAHGSGTKQNDKHETAAFRSLGDHAYR TPVSSIKMGVHSLGAIISIACVLMAESVIPPTANLHDDPEPECDLDYVPIITARE ARLDRLVSGSGFGGFSQAMVLTPRERNTA" /db_xref="GI:12744821" /gene="simA2" /db_xref="GI:12744821" /gene="simA2" /note="involved in polyketide formation" /codon_start=1 /transl_table=11 /product="putative ketosynthase SimA2" /protein_id="AAK06785.1" /db_xref="GI:12744821" /translation="MTTSVVVTGLGVAAPNGLTADYWAATREGSGIGRVTFRDPDSQ YPSRLAGEVPEFQVADHLPRLPQTDHMTLRALVSADWALQDAGIRPEELPEYAAVG VTASAGGPEFQNEHLRSLKSGVQVSAVQYFAWVAVNTQGISIRHGLRGPVGVV SDOAGLALQARQIRRGSOVTVSGGIDASICSWGAOLTSGRMSTDDDPATAYR PFDGAAGHPVEGEGALLVEAGAAARAGVYGRIGAGYTFPDKPSGREGRLR ALELATDAGLADPADIVFADAAVDPDLRIEARALVEVFGRLPVTAPKPSWTGRL NSGAGSLDVATALLAIRDGVPIPTTINVTAAQDDYELDVLVTAQARSRLRSALVVARHG GFNSALVVDQAA" /db_xref="GI:12744822" /gene="simA3" /db_xref="GI:12744822" /gene="simA3" /note="involved in polyketide formation" /codon_start=1 /transl_table=11 /product="putative acyl carrier protein SimA3" /protein_id="AAK06786.1" /db_xref="GI:12744822" /translation="MQTTEFTLDDLRILRDGAGASEVDVLEGGDIIDVAFEDLGYESI ALLETCSRIERYGISLDDDAVSEAPTPTPALVAIVNTHLIAVAG" /db_xref="GI:12744823" /gene="simA6" /db_xref="GI:12744823" /gene="simA6" /note="involved in polyketide formation" /codon_start=1 /transl_table=11 /product="putative ketoreductase SimA6" /protein_id="AAK06787.1" /db_xref="GI:12744823" /translation="MFGQKQVALVTGATSGIGLAVVRALAAQGHQVFPFCARNAEKVA VTIKELRGDLVDGTSVFMVTREVLTAGGMREKSSGRIVNIAGTAGQGVVLGAPYSAS TKELWADVTETLTSVFMVTREVLTAGGMREKSSGRIVNIAGTAGQGVVLGAPYSAS KXGVGPTKALGNELAPTGITVNAVCPGVETPMABERVQGVAAAGTSEALIEKFKQ 6810. .7757 /gene="simA5" /db_xref="GI:12744824" /gene="simA5" /note="involved in polyketide formation" /codon_start=1 /transl_table=11	gene	/product="putative cyclase/dehydrase SimA5" /protein_id="AAK06788.1" /db_xref="GI:12744824.1" /translation="MSDSGISHVEHTIEVDAPATDVYRLLAENVWNPRIFGPTIHVDH VERGARTERIRIATANGAKSWTSRRELDPEALIDFRQEVSAAPPVAAWGADIIIEP VSEGARVRLLDHRAIDDAEDLAWIEQAVDRSRSELAALKANAEVVGAADELML SFQDSVRIDGSAKDVYDNEADRWELPHVAEVELTEOTPLQVLTWETRAKDGST HTSSVRVCFPHERIVKQTVLPALMTLHTCYWLEEDGEGTVATSQHTVIINTANIR AILGESAGVAERQIVQTALSTNSRATLGHAKDYAERD" /db_xref="GI:12744825" /gene="simA8" /db_xref="GI:12744825" /gene="simA8" /note="involved in polyketide formation" /codon_start=1 /transl_table=11 /product="putative oxygenase SimA8" /protein_id="AAK06789.1" /db_xref="GI:12744825" /translation="MTGQPTDVLVTGPDPLMLAGELLGGARVVVAERLTPTTTSR ARRCMPTMEILDQRLGLRGEVPGDGTGHFGGIPRLDGGQSPYAGLWKVQSRVE EILGAWAIGLGAENVRRGRELTRREDGAVVEABLSCEGDTRELLRARIYVCGDGENST VRLAGFEFAGSAASRELIRADVIGIDVNNRRFERLPGLGLAIARRRPDGITRVMVHEF GRATPROAEPDFDEMARTWQVIGEDISHTGTPVINSFGDASRQVTEYRRGRILLAG DAAHAQMPVGGQALNIGLQDAFNLCWLAARICGWAPPGGLDLSYHERRHAGVSRVLSG IAQAQLLLGPEVEALRSVVRGLGLDAPVFPFGHRRHRLSSTPPTRVHHPHLLGA RLPDAALDIAGATAWALLRAGTGLGLLDVAPEAPGHVDLRAVAGWADRVRVVAARP RAGAVAPTTGAVLVRPDGHVVWADGGRLRAALVQWFGSADRT" /db_xref="GI:12744826" /gene="simA9" /db_xref="GI:12744826" /gene="simA9" /note="involved in polyketide formation" /codon_start=1 /transl_table=11 /product="putative reductase SimA9" /protein_id="AAK06790.1" /db_xref="GI:12744826" /translation="MKRLAGKATLVGSSRGIGRATLRLAREGATVAVHYADNADA KATLAAIEBEGRAFVRAELGVPGDVTDLFQGLTAGREATKEELNLVNNAGVWG GVKPEDLTPEIFDQLIAVNAKAPLQTVLSIMPDGGRINIGTGLLRYANPDEVAY SMKCALDQISLHPARLLPRNITVNTVSPGTTNNGSPITFDVPEAREFANMANAFKRI APEDVGDVAFPLASDDARWITGTFTVDASGGTLLGG" /db_xref="GI:12744827" /gene="simA10" /db_xref="GI:12744827" /gene="simA10" /note="involved in polyketide formation" /codon_start=1 /transl_table=11 /product="putative reductase SimA10" /protein_id="AAK06791.1" /db_xref="GI:12744827" /translation="MSSARLRLAVLSTREGRLCPTVANWFAGQGHGGFDLDLVD PYGLDIPVAPSEPEPPAQVVEIGORLAADAFVITPEYNHSPSAPLKSILDMHYVQWR AKPVGFVSYGGLAGGLRAVEQLVFAELHAATVTRDTSFHNAQLFDENGDRDPEG CNTAAKTLMDQLHWWGRALRQARLERYQG" /db_xref="GI:12744828" /gene="simA11" /db_xref="GI:12744828" /gene="simA11" /note="involved in polyketide formation" /codon_start=1	gene	Query Match 7.2%; Score 314.4; DB 1; Length 34869; Best Local Similarity 62.0%; Pred. No. 6.6e-19; Matches 517; Conservative 0; Mismatches 311; Indels 6; Gaps 1; QY 2977 CCGGCGGCGGCGCGCGGTAGTCTGCCCTCGGTACCGAAGCGTCCGGGTGAGAAACGTT 3036 Db 10131 CAGGTCCGCGCGCGGTGCGGATCGGCGCGCGCGCGGTGAGCGCGCGCGCGGTG 10072 QY 3037 CCGCCTGTGGCATCGACGCTACTGGCGCGGTGATCCACCGTGAGTCTCGGAGGCCAGAAAG 3096
------	---	------	--	------	--

[illegible]

Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Unwin, L., Whitehead, S., Barrell, B. G. and Maskell, D. J.

Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*

Unpublished

2 (bases 1 to 343243)

Direct Submission

Submitted (06-AUG-2003) Submitted on behalf of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: ms5@sanger.ac.uk

Location/Qualifiers

1. 343243

/organism="Bordetella pertussis Tohamia I"

/mol_type="genomic DNA"

/strain="Tohamia I"

/db_xref="taxon:257313"

194..2230

/gene="tktA"

/locus tag="BP0999"

/note="synonym: tkt"

194..2230

/gene="tktA"

/locus tag="BP0999"

/EC number="2.2.1.1"

/note="Similar to Escherichia coli transketolase 1 TktA Tkt or B2935 SW:TKY1_ECOLI (P27302) (663 aa) fasta score: E() : 5.7e-137, 63.43% id in 670 aa, and to Pasteurella multocida transketolase 1 TktA or Pml242 SW:TKY1_PASMU (P75927) (668 aa) fasta scores: E() : 3e-158, 61.504% id 678 aa"

/codon_start=1

/transl_table=1

/product="transketolase 1"

/protein_id="CAE41300.1"

/db_xref="GI:33571794"

/db_xref="GOA:Q7VZC0"

/db_xref="InterPro:IPR005474"

/db_xref="InterPro:IPR005475"

/db_xref="InterPro:IPR005476"

/db_xref="InterPro:IPR005478"

/db_xref="UniProt/TREMBL:Q7VZC0"

/translation="MSNPRTAPKLADAIRVLMDAVQKANSGHFGAPMGMAETGAQWLGHNRPDPANWNRDFVLSNGHSGMLIYALHLTGVDLPLEELKNFRQLHSRCHGEVGIPTGVETTTGLPGQLANAVGMAELAAEFNAGHTVVDHHTVVFGLGCLMEGISHEACSLAGTKLSKLVALYDDNGISIDGHVEHFADDTATEFEGVGNNIVGDGHDAVAADVAQKQAEKSEKTLIVCRVIVKGSNPMAGTHNVHGAPAGNDIEITRALNGSAGFEVPEVDIIYAQWDSRPAGALAAQWEDKFAQYAAQYPAEAAEFTGSNGELPGFPAARAEHFLQATVAKGETVATKASQFAIAELAGLLPEMLGGSADLTGSNDWKVGPFAARFGHGHGRIHNYGVREFGMAAINNGIALHGGYLPFGGTFLLTFSDDYNAIRMAALMKORVVHVHFTDSIIGLGEDGPTHQSIETHAASLILPNSVVRPCDVTVEAANGVAATPRASIGMDLHDGPTALLSRQNPFPVARSLEATQKAIVRGCVPLRDAAHRAAIATGSEVALEIALADGUAIGVVRVSMFSTDFDQDAEWKKEVLPAGMPAVEAGVATFWPHKYVLEGAVGVIDIRYGESAPAGALFKFGLGTAEKVAEAVQVL"

227..2203

/gene="tktA"

/locus tag="BP0999"

/note="HMMfam hit to PF00456, Transketolase, thiamine diphosphate binding domain"

239..301

/gene="tktA"

/locus tag="BP0999"

/note="ScanRegExp hit to PS00801, Transketolase signature"

2248..3258

/gene="gap"

/locus tag="BP1000"

/note="synonym: hexC"

2248..3258

/gene="gap"

/locus tag="BP1000"

/EC number="1.2.1.12"

feature

feature

```
/note="Similar to Pseudomonas aeruginosa glycerolaldehyde
3-phosphate dehydrogenase Gap or HexC or paf195
SWALL:G3P_PSEAE (SWALL:P27726) (334 aa) fasta scores: E():
9.1e-84, 66.07% id in 336 aa, and to Alcaligenes eutrophus
glycerolaldehyde 3-phosphate dehydrogenase, plasmid cbbgp
SWALL:G3PP_ALCEU (SWALL:P50322) (336 aa) fasta scores:
E(): 1.1e-107, 84.52% id in 336 aa"
/codon_start=1
/transl_table=11
/product="glycerolaldehyde-3-phosphate dehydrogenase"
/protein_id="CAE41301.1"
/db_xref="GI:33571795"
/db_xref="GOA:Q7VZB9"
/db_xref="InterPro:IPR000173"
/db_xref="UniProt/TREMBL:Q7VZB9"
/transl_table="MTIRVINGYGRIRNLRHAYHKKHDIIEVAINDLGDPKTN
AHLTFPTAHGKPGCTVGDGVVNGDKIRVLNARNPAELPMWELGVGVVLETCGF
FTSKKAGNHLGGAKVLIISAPGKDVDAITVFGVNGQNLKAEHTVVISNACTINCL
APLVKPLHDKLGVETGLMTTHAYTNDQVLTIDVHEDLRKARSATSMIPKTHGAHA
VGLVPELNGKLDGFAIRVPTINVSLLDLSFVAKGDTTVEEVNSILOAAEGLKDIIL
TYNTEPLVSIIDFNHPASSNFDTALTIVSGKLKVKVSSVYDNEWGFSNRMLDTTVALMS
AK"
misc_feature
2251. .3252
/gene="pgk"
/locus_tag="BP1000"
/note="HMMPfam hit to PF00044, Glycerolaldehyde 3-phosphate
dehydrogenase, NAD binding domain"
2704. .2727
/gene="gap"
/locus_tag="BP1000"
/note="ScanRegexp hit to PS00071, Glycerolaldehyde
3-phosphate dehydrogenase active site. Confirmed by
InterPro eMOTIF pattern match."
3376. .4569
/gene="pgk"
/locus_tag="BP1001"
3376. .4569
/gene="pgk"
/locus_tag="BP1001"
3376. .4569
/gene="pgk"
/locus_tag="BP1001"
/EC_number="2.7.2.3"
/note="Similar to Escherichia coli phosphoglycerate kinase
Pgk or B2926 SW:PGK_ECOLI (P11665) (386 aa) fasta scores:
E(): 6.3e-83, 64.042% id in 381 aa, and to Alcaligenes
eutrophus phosphoglycerate kinase, chromosomal cbbkc
SW:PGK_ALCEU (P50319) (413 aa) fasta scores: E():
1.8e-103, 75.381% id in 394 aa"
/codon_start=1
/transl_table=11
/product="phosphoglycerate kinase"
/protein_id="CAE41302.1"
/db_xref="GI:33571796"
/db_xref="GOA:Q7VZB8"
/db_xref="InterPro:IPR001576"
/db_xref="UniProt/Swiss-Prot:Q7VZB8"
/transl_table="MSNVNTLSALAKSGKRVFIRADLNVPFDDAGRISBETRIIR
ASVPGRLDALDAVMVTHSLGRPKEGALTEADSLAPVQRLSELGNOVLVDPDW
DGVSVRPGVLLLENNGVNGKDDDEGLSRKWAALCDVYVNDAPGTAHRAEDTHGI
ARFAPVACGPILEELDALGRALHDPKRPPIVAIVGGSKVSTKLSILOSLADKVDOLV
VGGGTANTPMLAAGPIGKSLAEPQSQVQARRAVIEIMKRGARVPIPTDVVCAK3FGA
DAAATVTKAADVAEDMILDIGPQTAQRLADILKTAGTIVWNGPVGVEFDQFAHGT
VVARAIADSAFSGIAGGGDTLAAIAKYGIADQGTGYISTGGGAFLEFLGKALPAVAL
QARA"
misc_feature
3421. .4563
/gene="pgk"
/locus_tag="BP1001"
/note="HMMPfam hit to PF00162, Phosphoglycerate kinase"
3433. .3465
/gene="pgk"
/locus_tag="BP1001"
/note="ScanRegexp hit to PS00111, Phosphoglycerate kinase
signature. Confirmed by InterPro eMOTIF pattern match."
4066. .4089
misc_feature
```

```
/gene="pgk"
/locus_tag="BP1001"
/note="ScanRegexp hit to PS00017, ATP/GTP-binding site
motif A (P-loop)."
complement(4599. .5372)
/gene="fabI"
/locus_tag="BP1002"
/note="synonym: envM"
complement(4599. .5372)
/EC_number="1.3.1.9"
/locus_tag="BP1002"
/note="Similar to Escherichia coli, and
enoyl-l-acyl-carrier-protein] reductase [NADH] FabI or EnvM
or B1288 or Z2512 or Ecs1861 SW:FABI_SCOLI (P29132) (261
aa) fasta scores: E(): 1.1e-35, 41.200% id in 250 aa, and
to Rhizobium meliloti putative enoyl-acyl-carrier-protein
reductase NADH TR:CAC41683 (EMBL:AL591783) (268 aa) fasta
scores: E(): 3e-40, 44.578% id in 249 aa"
/codon_start=1
/transl_table=11
/product="enoyl-l-acyl-carrier-protein] reductase [NADH]"
/protein_id="CAE41303.1"
/db_xref="GI:33571797"
/db_xref="GOA:Q7VZB7"
/db_xref="InterPro:IPR002198"
/db_xref="UniProt/TREMBL:Q7VZB7"
/transl_table="MNAAPLQGRRLVGTIANQDSIANGCARAFALGAELAVTYLN
DKARVHVPLARSIDAALLPLDLRVQDLGSVFQIGARWGLSLDFVHMSIAYPRDD
LRGVVDCSRAGFLQAMDISCWSFIRMAHPLMKGTLFCMTYYSQMVVHYNM
MGPVKAALSASTRYLAAELGPGQIRVHAISPGLKTRAASGITGFDALLDRAQAKAPA
RSLVGDIDVDEATLAWLATDAARRITGQTLIYIDGGTHIID"
complement(4626. .4718)
/gene="fabI"
/locus_tag="BP1002"
misc_feature
Query Match 7.0%; Score 305.2; DB 1; Length 343243;
Best Local Similarity 58.2%; Pred. No. 1.8e-18; Indels 6; Gaps 2;
Matches 576; Conservative 0; Mismatches 408;
QY 3011 GAAGCGTGGCGGGTCAGAGAAATCGTTCCGCCCTGTGGCATCGACGTACTGCGCGGTGATCC 3070
DB 185368 GGATCGACGGCGGCTACAGCGCTGCGCCCGCTGGCGTGGCGGTGAGGCGCTGCCCGGTGATCC 185309
QY 3071 ACCGTGAGTGTGCGGAGCGCAGAAAGCCACACCGTGGCGGATGTCTCGCGGTCTGCCCA 3130
DB 185308 AGCGCGCTGTGTCGAGGCGCAGAGGCGCACGATGTCCGCGATGTCTCGCGGTGCGCCA 185249
QY 3131 TCGCGTTGAACACGAGTGTGGCGCGCTGCGCGCGCGCGCTCGGGGTCTCGCGCGCGC 3190
DB 185248 CGCGCGCGAGCGCGATGTGCGCGCGCAG---CGCGTGGCGCATGGCGGGATCGCGGCGC 185192
QY 3191 GTGCGTTATGTCTCTCCAGAAACCGCGCGCCACCGCGTTCGACCGGTGATCCCGCGTT 3250
DB 185191 GCGTGTTCATGTCTGCGTGGCGTGGCGCGGCGAGCAGCGCATTCAGCGGTGATCGCGCG 185132
QY 3251 CCCCAGTTGCTGCGCAGGCGAGCGTGAAGCGTGTTCACCGCACCCCTTGGTTCATCGCGT 3310
DB 185131 CGCCAGGTGCTGCGCGCAGCAACAGGTCAGCGCTCCAGGCGCGCTTGGCGCGCGCT 185072
QY 3311 ATCCGATGACTCGGGGAACGCGCGCGGTGCGCGCAGACCGAGATGTTGATGATCCGCC 3370
DB 185071 AGACGCTCATTTTGGGGTAGCGCGCGCGCTGCCATCCAGCAGAGTTCAGATCGCGC 185012
QY 3371 CGCGGTGCGCGAGTCGTTTTCAGTCTGCTGCGACAGGAAACGCGGTGCGCGGACGCTTGA 3430
DB 185011 CGCGGTGCGCGCAGCAGCGCAGCGTGTGATCAGGAAGAACGCGGTTTTCAGTTGA 184952
QY 3431 CGCGGACCAAGTCGTCGAGACCTCTCTCGGTGACTTCCTCGTTCGTCGCGAGCCGCTGA 3490
DB 184951 CTGATGATCGGTTCGAGTCTCTCGGGGAGACCTCTCGGGCAGCGCGCGCGAGGCTTA 184892
QY 3491 CGCGCGCGTTGTTTACCAGGATGTGAACTCGGGGGGCGACTCCGAACCTCGCCATCCCG 3550
```


Db 317524 GGCCTGATCCGGGGCTGGCGCGGTGACCAAGGCGGTCTGCTGGCGCCAGGG 317574

Search completed: January 18, 2006, 16:18:34
Job time : 21080 secs